GREQAM
Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n° 2008-28

HIERARCHICAL HIDDEN MARKOV STRUCTURE FOR DYNAMIC CORRELATIONS: THE HIERARCHICAL RSDC MODEL.

Philippe CHARLOT
Vêlayoudom MARIMOUTOU

May 2008

Hierarchical Hidden Markov Structure for Dynamic Correlations: the Hierarchical RSDC Model.

Philippe CHARLOT[†], Vêlayoudom MARIMOUTOU[‡]

GREQAM & Université de la Méditerranée
Centre de la Vieille Charité
2, rue de la Charité
13236 Marseille cedex 02, France
[†] charlal13@yahoo.fr, [‡] Velayoudom.Marimoutou@cnrs-dir.fr

May 27, 2008

Abstract

This paper presents a new multivariate GARCH model with time-varying conditional correlation structure which is a generalization of the Regime Switching Dynamic Correlation (RSDC) of Pelletier (2006). This model, which we name Hierarchical RSDC, is building with the hierarchical generalization of the hidden Markov model introduced by Fine *et al.* (1998). This can be viewed graphically as a tree-structure with different types of states. The first are called production states and they can emit observations, as in the classical Markov-Switching approach. The second are called abstract states. They can't emit observations but establish vertical and horizontal probabilities that define the dynamic of the hidden hierarchical structure. The main gain of this approach compared to the classical Markov-Switching model is to increase the granularity of the regimes. Our model is also compared to the new Double Smooth Transition Conditional Correlation GARCH model (DSTCC), a STAR approach for dynamic correlations proposed by Silvennoinen and Teräsvirta (2007). The reason is that under certain assumptions, the DSTCC and our model represent two classical competing approaches to modeling regime switching. We also perform Monte-Carlo simulations and we apply the model to two empirical applications studying the conditional correlations of selected stock returns. Results show that the Hierarchical RSDC provides a good measure of the correlations and also has an interesting explanatory power.

Keywords: Multivariate GARCH; Dynamic correlations; Regime switching; Markov chain; Hidden Markov models; Hierarchical Hidden Markov models.

JEL Classification: C32, C51, G1, G0.

1 Introduction

Since the seminal papers of Engle (2002) and Engle and Sheppard (2001), the study of multivariate GARCH with dynamic correlations has given rise to many extensions and developments. This has been motivate by several empirical studies about stock market behaviors (like Longin and Solnik's papers (1995, 1996 and 2001)). They show that the hypothesis of constant correlations (CCC model of Bollerslev (1990)) is not realistic. However, if the assumption of dynamic correlations is now widely accepted in the literature, the question of the form of the dynamic is still an open problem; see Bauwens *et al.* (2006) or Silvennoinen and Teräsvirta (2007) for a recent survey.

The purpose of this paper is to present a new multivariate GARCH model with dynamic correlations. We propose a regime switching model that is part of the Markov-Switching class. It is a generalization of the RSDC model of Pelletier (2006). This model is halfway between CCC of Bollerslev (1990) and DCC of Engle and Sheppard (2001). Correlations are constant within regime, but vary from one regime to another and the transition between the different regimes are performed by a Markov chain.

Silvennoinen and Teräsvirta (2005) have proposed a model with smooth transition between regime for correlations (STCC model), which can be seen, under certain assumptions, as the competitive STAR approach of Pelletier's model. The STCC impose the correlations to varying between two matrix of constant correlations. The transition between this two extreme matrix is governed by a conditional logistic function. Recently, they build an extension of this model, the Double-STCC (see Silvennoinen and Teräsvirta (2007)) in which conditional correlations vary across four matrix of constant correlations through two logistic functions. Our new model, called Hierarchical-RSDC (HRSDC) can be seen as a Markov-Switching version of the DSTCC. In this new model, correlations vary between four correlation matrix constant in time but the transition from one to another is determined by hierarchical hidden Markov structure. The originality of this structure is to establish a hierarchy between the hidden states in order to increase the granularity of the regime. This hierarchical hidden structure was first developed by Fine *et al.* (1998) for handwriting recognition. In our context, this specific structure will allow us to bring out a finer definition of regimes that do not appear with the classical Markov-Switching approach.

The paper is organized as follows. In section 2, we briefly give a review of developments of dynamic correlations models. We also present a detailed overview of regime switching model, including STAR and Markov-switching approach, apply to correlation modeling. The Hierarchical-RSDC model is introduced in section 3.1.2. Section 4 presents results of Monte-Carlo simulations and two empirical applications and compare the HRSDC with the DSTCC and DCC models. Section 5 presents concluding remarks and expose some directions for future research.

2 DCC-MGARCH and Regime Switching models

The general framework of multivariate GARCH models with dynamic correlations assumes that a the stochastic process $\mathbf{r}_t$ followed by the observations of size $(K \times T)$ is defined by :

$$\mathbf{r}_t | \mathscr{F}_{t-1} \sim \mathscr{L}(0, H_t) \tag{2.1}$$

where $\mathscr{L}$ is a parametric distribution function with mean equal to zero and conditional variance H_t follows :

$$H_t = D_t R_t D_t \tag{2.2}$$

with $D_t = \text{diag}(h_{1,t}^{1/2}, ..., h_{K,t}^{1/2})$, a diagonal matrix composed of the standard deviation of the K univariate series, obtained for example with a GARCH filtering. This filtering permit to calculate the standardized residuals, which are expressed as $\epsilon_t = D_t^{-1}\mathbf{r}_t$. The expectation of these standardized residuals gives the conditional correlations:

$$\mathbb{E}_{t-1}[\epsilon_t \epsilon_t'] = D_t^{-1} H_t D_t^{-1} = R_t$$

The first multivariate volatility model to take into account the conditional correlations was the CCC of Bollerslev (1990). He adopts the simplest dynamic ever by taking the conditional correlations constant through time, i.e. $R_t = R \ \forall t$. The correlations matrix is then equal to the unconditional correlation matrix. However, several empirical studies show that correlations among assets are not constant. These results imply that the CCC seems too restrictive and give rise to an extensive literature about the process follows by the conditional correlations.

Tse and Tsui (2002) suggested to express the correlations with an autoregressive form (DCC_T model):

$$R_t = (1 - \theta_1 - \theta_2)R + \theta_2 R_{t-1} + \theta_2 \Psi_{t-1}, \ \theta_1 + \theta_2 \leq 1 \tag{2.3}$$

where R is the unconditional correlation matrix while the elements of the correlation matrix Ψ_{t-1} are defined as :

$$\psi_{ij,t-1} = \epsilon_{i,t-1}\epsilon_{j,t-1} / \sqrt{\epsilon_{i,t-1}^2 \epsilon_{j,t-1}^2} \text{ with } 1 \leq i \leq j \leq K$$

In the same way, Engle and Sheppard (2001) adopt an autoregressive formulation for R_t (DCC_{ES} model) with:

$$R_t = \text{diag}\{V_t\}^{-1/2} V_t \text{diag}\{V_t\}^{-1/2} \tag{2.4}$$

where V_t is the conditional covariance matrix such as:

$$V_t = (1 - \alpha - \beta)\overline{V} + \alpha \epsilon_{i,t-k}\epsilon_{j,t-k}' + \beta V_{t-l} \tag{2.5}$$

with $\overline{V}$ the unconditional covariance matrix. These initial formulations define the basic framework of dynamic correlations models and represent the starting point of many extensions that we will briefly recall (see also figure 1 for a mapping of these extensions).

Combining covariances for correlations. Engle (2002) generalize the model DCC_E by writing the equation 2 as:

$$V_t = [\iota\iota' - A - B] \odot \overline{V} + A \odot \epsilon_{t-1}\epsilon'_{t-1} + B \odot V_{t-1} \tag{2.6}$$

in which A and B are two symmetric matrix of size $K \times K$ and $\odot$ is the Hadamard product. This specification is a less restrictive than the previous. It has the disadvantage of having more parameters, but the advantage of not imposing the constraint of equal dynamics for all correlations. In order to reduce the explosive number of parameters for large correlation matrix, Hafner and Franses (2003) reach to write a parsimonious version of equation 2:

$$V_t = \bar{V}(1 - \bar{\alpha}^2 - \bar{\beta}^2) + \alpha\alpha' \odot \varepsilon_{t-1}\varepsilon'_{t-1} + \beta\beta' \odot V_{t-1} \tag{2.7}$$

where α and β are vectors of size $K \times 1$ rather than matrix, $\bar{\alpha} = (1/K)\sum_{i=1}^{N} \alpha_i$ and $\bar{\beta} = (1/K)\sum_{i=1}^{N} \beta_i$. This formulation (called GDCC) encompasses the generalization of Engle (2002). The parameter α_i can be seen as an indicator of the sensibility of the i^{th} series towards the correlations of the innovations. In other words, large α_i mean that correlations of the i^{th} series are sensitives to the correlations of the innovations. The extreme configuration where $\alpha_i = 0$ means that fluctuations of $r_{ij,t}$ are perfectly correlated with the ones of $q_{jj,t}$. Engle Cappielo and Sheppard (2006) propose another extension of the specification defined by equation 2 whose goal is to capture asymmetry in correlation dynamics. It consists of introducing a indicator function as follow:

$$V_t = (\bar{V} - A'\bar{V}A - B'\bar{V}B - G'\bar{N}G) + A'\epsilon_{t-1}\epsilon'_{t-1}A + B'V_{t-1}B + G'n_{t-1}n'_{t-1}G \tag{2.8}$$

with A, B and G diagonal matrix. The indicator n_t verify: $n_t = \mathbb{1}_{\{\epsilon_t < 0\}} \odot \epsilon_t$ and allows to calculate $\bar{N} = T^{-1}\sum_{t=1}^{T} n_t n'_t$. As we have seen, the Holy Grail of multivariate DCC-GARCH is to have the most flexibly specification. Several authors have tried to find a process which can express different dynamics at once. In this spirit, Billio *et al.* (2003a) defined block diagonal matrix in the equation 2 with their Block-Diagonal-DCC model. Nevertheless, this model seems to be hard to implement. That's why Billio *et al.* (2003b), with the Flexible-DCC, improved this idea by writing partitioned vectors:

$$V_t = cc' \odot \overline{V} + \alpha\alpha' \odot \epsilon_{t-1}\epsilon'_{t-1} + \beta\beta' \odot V_{t-1} \tag{2.9}$$

where α, β and c are the partitioned vectors. Finally, a few improvements allow this authors to present the Quadratic-Flexible-DCC which assume that:

$$V_t = C'\overline{V}C + A'\epsilon_{t-1}\epsilon'_{t-1}A + B' \odot V_{t-1}B \tag{2.10}$$

with A, B and C are defined as symmetric matrix. This model considers the Flexible-DCC as a special case where A, B and C are partitioned symmetric matrix.

On the other hand, Audrino and Trojani (2004) proposed a tree structured DCC-dynamics. The tree is build on V_t and thresholds are defined with ϵ_{t-1} via a weighted function:

$$\bar{\rho} = \frac{1}{K(K-1)} \sum_{u \neq v} \epsilon_{u,t-1}\epsilon_{v,t-1}$$

The partition $\mathscr{P} = \{\mathscr{R}_1, ..., \mathscr{R}_k\}$ of the state space $G = \mathbb{R}^{K+1}$ on $(\mathbf{r}_{t-1}, \bar{\rho}_{-1})$ give the threshold function:

$$V_t = f^{\mathscr{P}}(\mathbf{r}_{t-1}, \epsilon_{t-1}, V_{t-1}) = \sum_{i=1}^{k} c_i[(1 - \phi_i - \lambda_i)\bar{V} + \phi_i\epsilon\epsilon' + \lambda_i V]\mathbb{1}_{\{(\mathbf{r},\bar{\rho}) \in \mathscr{R}_i\}} \tag{2.11}$$

with $c \in (0, 1]$, $\phi_i, \lambda_i \geq 0$, $\phi_i + \lambda_i < 1$ and $\bar{V} \in \mathbb{R}^{2K}$. The partitioned plane $(\mathbf{r}_{t-1}, \bar{\rho}_{-1})$ brings out various groups of covariances.

Non-parametric correlations. A semi-parametric approach is presented by Hafner van Dijk and Franses (2005). In this Semi-parametric DCC model, this is the covariance matrix that is estimated by a non-parametric way:

$$V_t(x) = \frac{\sum_{t=1}^{T} \epsilon_t \epsilon_t' K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)} \tag{2.12}$$

where x_t is a observable variable (e.g. market return or $x_t = t$), $K_h(\cdot) = (1/h)K(\cdot/h)$ a kernel function and h a bandwidth parameter. So, correlations are estimated with a modified Nadaraya-Watson estimator. We would further point out that the choice of the observable variable is tricky. First, this approach assumes that the conditional correlations depends on a exogenous variable. This variable will influence the estimated correlations by conditioning the kernel. Second, the authors indicates that the choice of the bandwidth highly influence the calculus of the correlations and suggests to use local bandwidths instead of a constant bandwidth.

In order to avoid the problem of sparse data bring up by Hafner van Dijk and Franses (2005), Feng (2006) develops a local estimator for the correlations using k-NN methods (k-nearest-neighbours). The idea of k-NN methods is to classify each observations with a particular decision rule. The goal is to find the k nearest observations to construct a bandwidth for these k ones. It can be done by considering the regressor $\mathbf{w}_{jt} = \mathbf{1}'\epsilon_{t-j}\epsilon_{t-j}\mathbf{1}$, where $\mathbf{1}$ is a vector of length K, $j = 1, ..., p$ and $t = p+1, .., T$. To obtain this moving bandwidth, it is necessary to consider two smoothing parameters: (i) an integer k such that $k \to \infty$ and $k/n \to 0$ when $n \to \infty$, (ii) choose b_0 be the half bandwidth for the re-scaled time such that $b_0 \to 0, nb_0 \to \infty$ and $(nb_0)^{-1}k \to 0$ when $n \to \infty$ with $t_0 = [n\tau]$[1] such that $t_0 > p$. Defining $k_1 = [nb_0]$ and $k_0 = 2k_1 + 1$, the bandwidth for $\mathbf{w}_j$ is chosen as:

1. calculate $n_1 = t_0 - k_1$ et $n_2 = t + k_1$
2. calculate $d_t = ||(\mathbf{w}_t - \mathbf{w})'(\mathbf{w}_t - \mathbf{w})||^{1/2}$
3. sort d_t according to an ascending order

The nonparametric estimator is builded with two kernel. The first $K_0(\cdot)$ is an univariate one with support $[-1; 1]$. The second, $K(\cdot)$, is p-dimensional. The proposed estimator can be written as:

$$V_\tau = \frac{\sum_{t=n_1}^{n_2} \epsilon_t \epsilon_t' K_0(\frac{\tau_t - \tau}{b_0}) K(\frac{\mathbf{w}_{1t} - \mathbf{w}_1}{b}, ..., \frac{\mathbf{w}_{pt} - \mathbf{w}_p}{b})}{\sum_{t=n_1}^{n_2} K_0(\frac{\tau_t - \tau}{b_0}) K(\frac{\mathbf{w}_{1t} - \mathbf{w}_1}{b}, ..., \frac{\mathbf{w}_{pt} - \mathbf{w}_p}{b})} \tag{2.13}$$

Feng points out that in practice, it seems appropriate to take $n_1 = t - k_1$ and $n_2 = t - 1$, i.e. to take a smoother involving only past observations.

Direct specification for correlations. Until now we have only discussed about specifications that were extensions of the basic model of Engle and Sheppard (2001). Recall that models calculate correlations from covariances. Others models specified directly the correlations.

Such is the case of Palandri (2006) with the Sequential-Conditional-Correlations model. His idea is to use a sequential Cholesky decomposition to obtain the correlation matrix. Correlations are separated between the conditional correlations and partial correlations as follow:

$$R_t = [\prod_{i=1}^{K-1} \prod_{j=i+1}^{K} K_{i,j,t}][\prod_{i=1}^{K-1} \prod_{j=i+1}^{K} K_{i,j,t}]' \tag{2.14}$$

The matrix $K_{i,j,t}$ is a lower triangular matrix with elements such as:

$$K_{i,j,t}[row, col] = \begin{cases} \rho_{i,j,t} \text{ if } row = j \text{ and } col = i \\ \left(1 - \rho_{i,j,t}^2\right)^{1/2} \text{ if } row = j \text{ and } col = j \\ \mathbf{I}_{[row,col]} \text{ otherwise} \end{cases}$$

[1] The operator $[x]$ denote the largest integer which is smaller than x.

where **I** is the identity matrix. Moreover, instead of taking a BEKK representation as usual, he builds an originally process based on the Fisher transform : $\chi_t = 0.5 \ln((1+\rho_t)/1-\rho_t)$. Correlations follow then a bivariate autoregressive process and become:

$$\chi_t = \omega + \sum_{h=1}^{p} \delta_h \chi_{t-h} + \sum_{i=1}^{q} (\theta_i + \beta_i d_{t-i}) \psi_{t-i}$$

with :

$$d_{t-i} = \begin{cases} 1 \text{ if } \epsilon_{i,t-1} < 0 \text{ and } \epsilon_{j,t-1} < 0 \\ 0 \text{ otherwise} \end{cases} \quad \text{and } \psi_t = \frac{1}{2} \ln(\frac{1+\phi_t}{1-\phi_t})$$

where ϕ_t is an evaluation of the *realized* correlations at time t, which can be computed by an exponential smoothing. Palandri also presents extensions of his model with the introduction of exogenous variables or dummy for asymmetry. If its implementation remains delicate, the advantage of the SCC model is to be able to calculate large correlations matrix very quickly through the sequential decomposition of R_t.

Finally, two others models have a direct specification for correlations. The goal of this two competing approaches is to modeling regime switching for the correlations. The first one is the STAR type model of Silvennoinen and Teräsvirta (2005). The second is from Pelletier (2006) and is builded in a Markov-Switching setup. Recall that this two main classes of models -STAR and Markov-Switching- are build on two ways to treat the problem. The STAR class assumes that the switch between the regimes is deterministic and determinate by a observable transition variable. The Markov-Switching class is quite different. It assumes that the process followed by the observation is governed by an underlying hidden stochastic process. To better understand our proposed model which generalizes the Markov-Switching of Pelletier, both approaches are described in more details in the following subsection.

2.1 (S)TAR approach

2.1.1 Theorical aspects

The Threshold Autoregressive (TAR) model of Tong (1980) was one of the first approach to model regime switching. It assumes that at each time t regime is determined by a threshold variable and a threshold value. A two regimes model will then be composed by two parallel sub-regimes where the transition from one to another depends of the position of the threshold variable against the threshold value. The SETAR model, for Self-Exciting TAR, is an direct extension in which the threshold variable is the lagged time series itself. For example, a basic 2-regime SETAR model is written as:

$$y_t = \Phi_1(L) y_t \mathbb{1}_{\{y_{t-d} > \gamma\}} + \Phi_2(L) y_t (1 - \mathbb{1}_{\{y_{t-d} > \gamma\}}) + \varepsilon_t \tag{2.15}$$

where $\mathbb{1}_{\{y_{t-d} > \gamma\}} = 1$ if $y_{t-d} > \gamma$ and 0 otherwise, $\Phi_1(L)$ and $\Phi_2(L)$ are two lagged polynomial of finite order and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$.

However, these early models are distinguished by a very sudden switch between regimes. Teräsvirta (1994) proposes another extension in order to avoid this problem with the Smooth Transition AR (STAR). His solution consists to replace the indicator function by a continuous function $G(y_{t-d}; \gamma, c)$ which vary smoothly and monotically from zero to one. This transition function depends on three variables. The first is the transition variable y_{t-d}; the two others are parameters which determined the localization (c) and the smoothness (γ). The STAR version of the process defined by equation 2.15 is then expressed as:

$$y_t = \Phi_1(L) y_t G(y_{t-d}; \gamma, c) + \Phi_2(L) y_t (1 - G(y_{t-d}; \gamma, c)) + \varepsilon_t \tag{2.16}$$

Along all possible functions, the most commonly-used are the logistic function (LSTAR model):

$$G(y_{t-d}; \gamma, c) = (1 + \exp(-\gamma \prod_{i=1}^{n} (y_{t-1} - c_i)))^{-1}, \; c_1 < ... < c_n, \; \gamma > 0 \tag{2.17}$$

and the exponential function (ESTAR model) :

$$G(y_{t-d}; \gamma, c) = 1 - \exp(-\gamma \prod_{i=1}^{n} (y_{t-1} - c_i)^2), \; c_1 < ... < c_n, \; \gamma > 0 \tag{2.18}$$

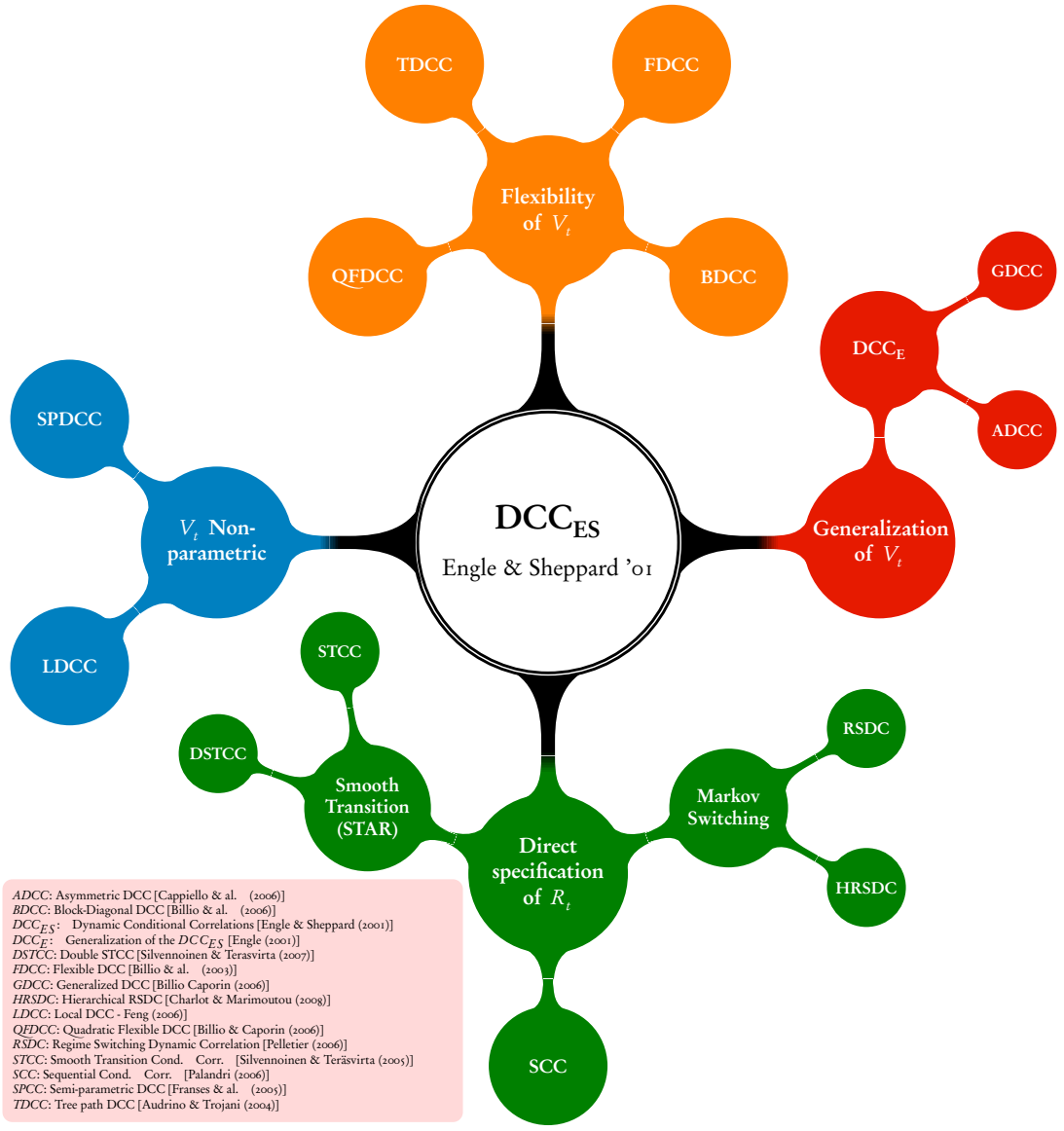

Figure 1: The different extensions of DCC model of Sheppard and Engle (2001). The model we propose is the RSDC hierarchical (HRSDC). As this mapping shows, it aims to be similar of the STAR approach of the DSTCC but in a Markov-Switching framework.

Figure 2 shows this two functions for various smoothness values. The main advantage of the STAR model is the effect of the smoothness parameter. It provides a strong transition when $\gamma \to +\infty$ and the STAR can be seens as a TAR. Consequently, when $\gamma \to 0$, the transition is very smooth and when $\gamma = 0$, the STAR performed as a linear model (see Terasvirta *et al.* (2000) for a complete survey, or Franses and van Dijk (2004)).

These methodologies, first developed for autoregressive processes, were quickly applied to GARCH models. Tong's approach has been taken over by Glosten *et al.* (1993) for the GJR-GARCH model, in which the conditional variance follows two different processes depending on the sign of error terms. In the same spirit, the model TGARCH developed by Rabemananjara and Zakoian (1993) specifies various regimes for the dynamics of the conditional standard deviation depending on the sign of error terms. Finally, Li and Li (1996) develop the Double-Threshold ARCH (DTARCH), which imposes a threshold structure for conditional variance and also for the conditional mean. Following the example of the latter model, Koutmos (1998) proposes a dual-threshold model in order to test the asymmetrical response of

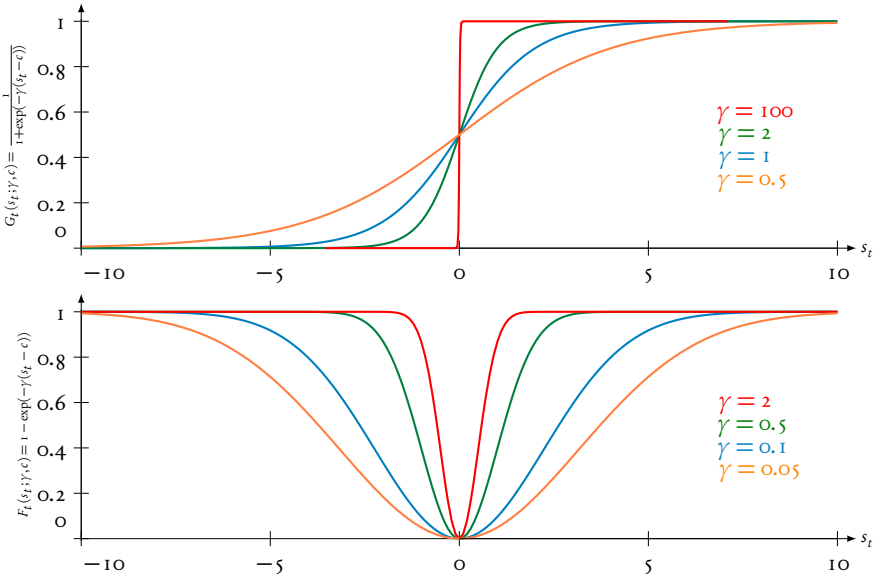

Figure 2: Examples of logistic and exponential functions with $c = 0$ and different values of γ.

mean and conditional variance with past information (asAR-TGARCH model). The STAR approach also results in various solutions capable of taking into account the asymmetry. Hagerud (1996) makes two STAR-GARCH models. The first type is LSTAR where the transition variable is based on the sign of residuals. The second model is based on a smoothing ESTAR and the transition variable is the absolute value of residuals. These two models are complemented by the analysis of Gonzales-Riviera (1998) with take as transition variable the asymmetry parameter of the variance. Lundberg and Teräsvirta (1999) build the STAR-STGARCH model in which the mean follows a STAR process and the conditional variance follows a STAR-GARCH. This approach extends two other nonlinear approaches that are GJR-GARCH and Quadratic GARCH of Santana (1995). Nam *et al.* (2002) apply a non-linear structure based on the STAR approach to the mean and conditional variance to capture the asymmetry (ANST-GARCH model). There is also the approach of Lanne and Saikkonen (2005), in which the transition variable is the conditional variance and whose goal is to model series with high persistence in the conditional variance. Finally, Medeiros and Veiga (2007) establish a general formulation with the Flexible Coefficient Smooth Transition GARCH model (FC-STGARCH), encompassing most of the models mentioned above.

2.1.2 STCC model

The construction of a STAR structure for the process followed by the correlations of a multivariate GARCH model was produced by Silvennoinen and Teräsvirta (2005) and Berben and Jansen (2005), leading to the Smooth Transition Conditional Correlation GARCH model (STCC-GARCH). The idea is to establish a conditional correlation matrix R_t as a combination of two extremes correlation matrix R_1 and R_2 constants over time. The transition from one to another is provided by the logistic function as defined by the equation 2.17. Specifically, the STCC specifies dynamic of the correlations as:

$$R_t = (1 - G_t)R_1 + G_t R_2 \text{ with } G_t = (1 + e^{-\gamma(s_t - c)})^{-1} \tag{2.19}$$

where γ is the smoothing parameter, c is setting the localization and s_t is the transition variable. To ensure the semi-positivity of the conditional matrix R_t, the matrix R_1 and R_2 must themselves be positive definite, via a Cholesky decomposition for example. The transition variable s_t can be both deterministic or stochastic, but its choice will remain crucial. If there is no real consensus to choice this variable, the study of empirical analysis using the STCC model shows two types of transition variable. It can be:

- a endogenous or exogenous variable. In Silvennoinen and Teräsvirta (2005), the authors apply the model STCC to study the correlations of assets from the daily S&P500. The transition variable is the value of the lagged returns of the squared S&P500. Their strategy is builded on the idea that a

strong difference value of the index at a given time represents a strong turbulence of the assets of the S&P500 to the next period.

- calendar time, i.e. $s_t = t/T$. This specification results in the terminology of Silvennoinen and Teräsvirta (2005) to the Time-Varying Smooth Transition Conditional Correlation (TVSTCC) model. That's the choice of Berben and Jansen (2005) in order to study correlations of equity returns.

Silvennoinen and Teräsvirta establish a strategy of inference for choosing the best transition variable. The idea is to find the most explicative variable controlling the dynamics of correlations. In this context, the transition variable varies depending on the purpose of the analysis. In the case where preliminary analysis did not reach to discriminate clearly this transition variable, the choice of calendar time seems to be the most consensual.

These authors extended the STCC model by introducing two transitions functions to evolve correlations between four correlation matrix constants over time. The conditional correlations process then takes the following form:

$$R_t = (1 - G_{1t})R_{(1)t} + G_{1t}R_{(2)t} \text{ avec } \begin{cases} R_{(i)t} = (1 - G_{2t})R_{(i1)} + G_{2t}R_{(i2)} \\ G_{it} = (1 + \exp(-\gamma(s_{it} - c_i)))^{-1} \end{cases}, i = 1, 2 \tag{2.20}$$

As for the STCC model, the transition variable can be of various kinds (endogenous or exogenous). We now focus on a special case of the DSTCC, in which transition functions are $s_{1t} = s_{2t} = t/T$. In this version, the correlations dynamic can be rewritten in a more illustrative way as:

$$R_t = (1 - G_{2t})\underbrace{((1 - G_{1t})R_{11} + G_{1t}R_{12})}_{R_{1t}} + G_{2t}\underbrace{((1 - G_{1t})R_{21} + G_{1t}R_{22})}_{R_{2t}} \tag{2.21}$$

In this formulation, the correlation matrix is a linear combination of two conditional matrix R_{1t} and R_{2t}. These two matrix are themselves a linear combination of two extreme matrix constants over time and semi-positive definite.

Estimation is done by maximizing the likelihood with the usual iterative optimization methods (i.e. Gradient), in one or two steps in accordance with the sample size (see Engle and Sheppard (2001)). However, as often in the regime switching models, the existence of numerous local maxima require to test many initial conditions.

2.2 Markov-Switching approach

2.2.1 Theorical aspects

The Markov Switching approach has its origins in the construction of mixtures models[2]. The objective of this models is to estimate a density assuming that the density is a finite mixture of densities, i.e. a convex combination of densities. Let $\{\mathscr{Y}_t\}_{t\in\mathbb{N}^*} = \{\mathbf{y}_1, ..., \mathbf{y}_T\} \subset \mathbb{R}^k$ T observations of length k with the PDF $f(\mathbf{y}_t)$ on $\mathbb{R}^k$. Then, the density of $\{\mathscr{Y}_t\}$ is given by:

$$f(\mathbf{y}_t) = \sum_{i=1}^{N} \pi_i f(\mathbf{y}_t; \phi_i) \tag{2.22}$$

where $f(\mathbf{y}_t; \phi_i)$ is a density of parameters ϕ_i. The parameter $\pi_i \in (0, 1)$ corresponds to the weight of the component i in the mixture. To ensure the function $f(\mathbf{y}_t)$ to be a probability density function, the positive sequence $(\pi_i)_{i\in\mathbb{N}^*}$ must then verify that $\sum_{i=1}^{N} \pi_i = 1$. Estimation of finite mixture model is particularly sensitive for the simple reason that we don't know the proportion of each of the components in the mixture model. More specifically, we may not affect the observations to the components of the mixture. An usual solution to this problem is then to reconsider the model in a incomplete data framework. This method combines to each vector $\mathbf{y}_t$ an indicator variable $\{\mathscr{S}_t\}_{t\in\mathbb{N}^*} = \{s_1, ..., s_T\} \subset \mathbb{R}$ such that $\mathbb{1}_{\{s_n=i\}} = 1$. The incomplete data model is so represented by the pair $\{\mathscr{Y}_t, \mathscr{S}_t\}_{t\in\mathbb{N}^*}$. The data are supposed to spread over the various components. The weight of a component can then be interpreted

[2] See, for example, McLachlan and Peel (2000) for a briefing and Böhning *et al.* (2007) for a survey of recent developments.

as the probability that an observation belongs to the component i : $\pi_i = \mathbb{P}[s_t = i]$. The model is built from an underlying structure composed of the latent indicator variables *iid* $\mathbb{1}_{\{s_n=i\}}$. The conditional probability and the joint probability are written as:

$$\mathbb{P}[\mathbf{y}_t|s_t] = \sum_{i=1}^{N} f(\mathbf{y}_t;\phi_i)\mathbb{1}_{\{s_n=i\}} \text{ et } \mathbb{P}[\mathbf{y}_t, s_t] = \sum_{i=1}^{N} \pi_i f(\mathbf{y}_t;\phi_i)\mathbb{1}_{\{s_n=i\}} \tag{2.23}$$

The mixture model can be seen as a latent variable model assuming the observed data is an indicator of belonging to a component. This approach can be enriched assuming that the indicator follows a hidden Markov chain, which allows us to introduce a temporal dimension on the indicator components.

Recall briefly the main definitions and properties needed on Markov chains[3]. Let $\mathcal{S}$ a countable set called *state space*, in which each $i \in \mathcal{S}$ are called *state*. A Markov chain of order k is defined by a sequence of random variables $(s_t)_{t\in\mathbb{N}}$ taking values in $\mathcal{S}$ such that for each $i_t \in \mathcal{S}$ and $t \geq 1$:

$$\mathbb{P}[s_{t+1} = i_{t+1}|s_t = i_t, ..., s_0 = i_0] = \mathbb{P}[s_{t+1} = i_{t+1}|s_t = i_t, ..., s_{t-k+1} = i_{t-k+1}] \tag{2.24}$$

Mixture models based on a Markov chain generally use a Markov chain of order one[4], which admits a simple intuitive interpretation: *the current state depends only on its previous state*. The dynamic is entirely based on the transition probability between states: $a_{ij} = \mathbb{P}[s_{t+1} = j|s_t = i]$. A Markov chain $(s_t)_{t\in\mathbb{N}}$ is said to be time-homogeneous if the probability of transition from one state to another are independent of t, i.e. $\mathbb{P}[s_{t+1} = j|s_t = i] = \mathbb{P}[s_1 = j|s_t = i]$. The transition probabilities of a homogeneous Markov chain with n states lead to a transition matrix:

$$\mathcal{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{2.25}$$

which verify $\sum_j a_{ij} = 1$ and $0 \leq a_{ij} \leq 1$. Any transition matrix can be represented by a directed valued graph whose summits represent the states of the chain. Figure 3 exhibits an graphical example of transitions between three states. Thus, assuming that the initial distribution of the states is π_t, we can easily obtain the distribution at $t+1$ with: $\pi_{t+1} = \pi_t \mathcal{A}$. The behavior of the transition matrix for the m next periods can be deduced with Chapman-Kolmogorov theorem:

$$\mathcal{A}^{(m)} = \begin{bmatrix} a_{11}^{(m)} & \cdots & a_{1n}^{(m)} \\ \vdots & \ddots & \vdots \\ a_{n1}^{(m)} & \cdots & a_{nn}^{(m)} \end{bmatrix} = \mathcal{A}^m \tag{2.26}$$

where $a_{ji}^{(m)}$ is the probability to move from state i at time 0 to state j at time m and for each $n, k \in \mathbb{N}$, $h = 1, ..., k-1$: $a_{ji}^{(k)} = \sum_{l\in S} a_{li}^{(h)} a_{lj}^{(k-n)}$. It is then possible to identify different types of states by noting that the space $\mathcal{S}$ can be partitioned. Let i and j two states of the chain, it is said that i*is accessible to* j, written as $i \to j$, if $\mathbb{P}[s_{t+k} = j|s_t = i] > 0$. Similarly, these two states *communicate*, i.e. $i \leftrightarrow j$, if $i \to j$ and $j \to i$. A state is called *recurrent* if $\mathbb{P}[\cup_{n=1}^{\infty}(s_n = i)|s_0 = i] = 1$ and called *transient* if $\mathbb{P}[\cup_{n=1}^{\infty}(s_n = i)|s_0 = i] < 1$[5]. A Markov chain in which all states communicate is called *irreducible*.

The basic idea of the Hidden Markov Model (HMM) is that the state of the system is not determined by observations, but through a hidden process which establish a relation between an observation and a state. Each observation y_t is linked to a random variable s_t following a Markov chain by a probability. The assumption that the random variables associated with observations follow a first order Markov chain implies a temporal dimension, since the probability of being in a state in $t+1$ depends on the state in t. The main difference with the finite mixture model deals on the indicator function, which here admits a temporal dependence. The random variable s_t generally follows (but not necessarily) a Markov chain

[3] For a more complete outline, we refer to Billingsley (1986, chap. 1.8) and Shiryaev (1996, chap. 1.12 and 8).
[4] $\Pr[S_{t+1} = s_{t+1}|S_t = s_t, ..., S_0 = s_0] = \Pr[S_{t+1} = s_{t+1}|S_t = s_t]$
[5] The quantity $\mathbb{P}[\cup_{n=1}^{\infty}(S_n = i)|S_0 = i]$ is the conditional probability to come back to state i given the chain started at state i.

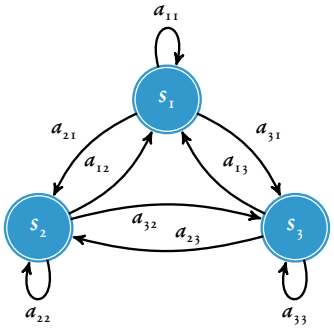

Figure 3: Basic structure of a first order Markov chain with three states.

of order one, but mostly complies with the conditions of homogeneity, ergodicity and irreducibility. Let N the number of states, an HMM is determined by the couple $\{\mathscr{Y}_t, \mathscr{S}_t\}_{t\in\mathbb{N}}$ and is defined by three elements: (i) a vector of initial distribution of size $(N \times 1)$: $\nu = P[s_1 = i] \ \forall i \in S$, (ii) a transition matrix $A = (a_{ij})$ of size $(N \times N)$ such that $a_{ij} = P[s_{t+1} = j | s_t = i] \ \forall i, j \in S$, (iii) a distribution of observations, independents conditionally to the hidden states, : $P[y_t | s_t = i] = f(y_t; \phi_i)$ where ϕ_i is the parameters of a parametric density family. In the discrete case, the joint probability followed by the hidden Markov model defined by the couple $\{\mathscr{Y}_t, \mathscr{S}_t\}_{t\in\mathbb{N}}$ is such as:

$$\mathbb{P}[\mathbf{y}_{1:T}, s_{1:T}] = \mathbb{P}[\mathbf{y}_{1:T} | s_{1:T}] \times \mathbb{P}[s_{1:T}] = \prod_{t=1}^{T} \mathbb{P}[\mathbf{y}_t | s_t] \times \prod_{t=1}^{T-1} \mathbb{P}[s_{t+1} | s_t] \times \mathbb{P}[s_1] \tag{2.27}$$

Partly introduced by Lindgren (1978) in economics and highly popularized by Hamilton (1989), the Markov-Switching approach is a slight variation of HMM[6] in which the conditional distribution of y_{t+1} depends not only on s_{t+1}, but also on y_t and/or y_{t-1}, etc. Nevertheless, the Markov-switching approach uses the same techniques for estimating than the HMM. The Markov-Switching version of the model described by the equation 2.15 is written as:

$$y_t = \Phi_{s_t}(L) y_t + \varepsilon_t \tag{2.28}$$

where $\Phi_{s_t}(L)$ corresponds of the specification of the regime fitting with s_t. Since the seminal paper of Hamilton (1989), this approach gave rise to a huge literature. We can quote, without being exhaustive, the work of Hamilton (1990) for the development of a filter to estimate the transition probabilities, as well as Krolzig (1997 and 2003) for a comprehensive extension of the original model to a multivariate framework. Kim (1994) also introduces a smoothing filter refining the calculation of posteriors probabilities.

The first ARCH model with Markov-switching (SWARCH) is proposed by Hamilton and Susmel (1994) while the first GARCH extension (GRS) appears in an analysis of the interest rates by Gray (1996). The innovation for the model developed by Gray lies in writing the conditional variance. Indeed, for SWARCH, each conditional variance depends not only on the regime, but on all history[7]. Whereas in the GRS, the conditional variance depends only on the current regime by the expectation of the lagged variance. Klaassen (2002) improves the latter model by broadening the spectrum of information for the lagged variance. Haas *et al.* (2006) suggest a slightly different approach with the MSG(N) model, which consists of estimating N GARCH models in parallel and switching directly from a variance to another. They then avoid the problem of path-dependency. It generalizes several Markov-switching GARCH approaches like the models proposed by Haas *et al.* (2004) and by Francq *et al.* (2001), which appear as special cases. Finally, Liu (2006) deepens the MSG(N) model by studying the conditions of his steady states with less restrictive assumptions than Haas *et al.* (2006).

[6] In fact, these two names are not yet standardized and a model Markov-switching is often called the HMM (see Cappé (2005, chap. 1.2).

[7] It is the well-known path-dependency problem.

2.2.2 RSDC model of Pelletier (2006)

The literature on Markov-Switching GARCH models in a multivariate framework is actually much less dense than in the univariate context. Lee and Yoder (2005) apply the Gray's method to build a bivariate Markov-switching BEKK model. The multivariate GARCH model of Pelletier (2006) is the first, and for now the only one, to adopt a Markov-Switching structure for the correlations process. By imposing constants correlations in each regime and establishing the switch of each other through a Markov chain of order one, this model falls between the CCC and the DCC. Formerly, the RSDC model assumes that the conditional correlation matrix has the form:

$$R_t = \sum_{n=1}^{N} \mathbb{1}_{\{s_n=i\}} R_n \tag{2.29}$$

where $\{s_t\}_{t\in\mathbb{N}}$ is a sequence follows a homogeneous first order Markov chain with N state. R_n is a conditional correlation matrix of size $K \times K$ where $R_n \neq R_{n'}$ for $n \neq n'$. The distribution of the process defined by $\{s_t\}_{t\in\mathbb{N}}$ is determined by a transition matrix written A. So that the conditional correlation matrix is indeed a correlation matrix[8], it is necessary to impose a PSD constraint on the matrix corresponding to the regimes. This can be a constraint Choleski decomposition, i.e. $R_n = C_n C_n'$ with C_n a lower triangular matrix and imposing that diagonal elements of C_n to be written as:

$$c_{j,j} = \sqrt{1 - \sum_{i=1}^{j-1} c_{j,i}^2} \tag{2.30}$$

Pelletier also proposes a more parsimonious specification of the RSDC whith the following expression:

$$R_t = \bar{Q}\lambda_{s_t} + \mathbf{I}_K(1 - \lambda_{s_t}) \tag{2.31}$$

With two states, the latter expression as being equal to the weighted average of two extreme states. The case $\lambda_{s_t} = 0$ is where the returns are not correlated while $\lambda_{s_t} = 1$ corresponds to the case where they are very correlated. This second formulation requires an identification constraint as follows:

$$\lambda_1 = 1 \text{ with } \lambda_1 > \lambda_2 > ... > \lambda_N \text{ or also } \max_{i\neq j} |R_{ij}| = 1 \text{ with } 1 > \lambda_1 > \lambda_2 > ... > \lambda_N \tag{2.32}$$

Estimation of the RSDC model is done via maximum likelihood, with a nuance according to the specification in question. For the first specification, corresponding to the equation 2.29, Pelletier favors the use of the EM algorithm. The second specification, define by equation 2.31, can be estimated through Gradient methods by exploiting the idea of correlation targeting of Engle et Mezrich and using Hamilton's filter. In that case, the correlations can be simply calculated as:

$$\mathbb{E}[R_t] = \bar{Q}\sum_{n=1}^{N} \lambda_n \pi_n + I_K \sum_{n=1}^{N} (1 - \lambda_n)\pi_n \tag{2.33}$$

3 Hierarchical Hidden Markov Structure for Dynamic Correlations

As we have seen in section 2, the DSTCC model of Silvennoinen and Teräsvirta (2007) is an expanded version of the STCC. Its feature is to express the conditional correlation matrix as a combination of two extreme matrix that are themselves dynamics. As the authors explain, at the beginning of the sample, the correlations defined by the equation 2.21 will vary between the two states $R11$ and $R12$, and then towards the end of the sample between R_{21} and R_{22}. Couples of matrix (R_{11}, R_{12}) and (R_{21}, R_{22}) define, according to the second transition variable, two major states bounded by two extreme dynamic matrix. We can say that the dynamic of the DSTCC works on two *primaries* regimes, themselves build under four *secondaries* regimes. The Markov-Switching approach of Pelletier (2006) doesn't consider the case of sub-regimes, unless, in our case, to take four states. However, a classical Markov-Switching model

[8] Recall that by definition, for R_t be is a correlation matrix, it must be symmetrical, PSD, have its diagonal elements equal to 1 and the non-diagonal elements must belong to the interval $[-1; 1]$.

will not reach to capture the existence of sub-regimes. The idea of the model that we present is that, unlike the basic Markov-Switching, it makes possible to establish a hierarchy between states that define sub-regimes. The aim is to show nuances induced by the existence of sub-regimes in the dynamic of the correlations. As a first step, we will present the class of hierarchical hidden Markov models, the structure which is build our model. In a second step, we will expose our model.

3.1 Structure

3.1.1 The Hierarchical Hidden Markov Model

The Hierarchical Hidden Markov Model (HHMM) has been proposed by Fine *et al.* (1998) in order to generalize the HMM model. The idea is to build a stochastic process with several levels by adopting a tree structure to obtain an interlacing of regimes. The hierarchy of the tree is built with *internal states*, which are abstract states (i.e. they do not produce observations). The internal states can lead so-called *emitting states*, which produce observations. Internal and emitting states can also lead to a third type of state which are called *exiting states*. The exiting states allow to quit a level of the tree. Each internal state produce a sub-HMM, which can also be itself a HHMM. In this framework, the classical HMM is a special case of the HHMM with only one level. The main advantage of the HHMM compared to the HMM is to improve the *granularity* of the regimes. The HHMM approach permits to break up a time series into several types of regimes. For example, a HHMM with two levels has two types of regimes: the firsts are called *primaries regimes* and the seconds *secondaries regimes*. The combination of the secondaries permits to deduce the primaries ones. This increased granularity allows us to bring out nuances in regimes which can not occur with the simplified structure of HMM.

Formally, a HHMM can be represented as the process $\{\mathscr{Y}_t, \mathscr{Q}_t\}_{t\in\mathbb{N}}$ with :

- $\{\mathscr{Y}_t\}_{t\in\mathbb{N}}$ is the process followed by the observations, which are supposed to be conditionally independent to the hidden states.
- $\{\mathscr{Q}_t\}_{t\in\mathbb{N}}$ is a homogeneous first order Markov chain. Each state of a HHMM q_i^d belongs to the set $\mathscr{Q} = \{\mathscr{S}, \mathscr{I}, \mathscr{E}\}$ where $\mathscr{S}$ is the set of emitting states, $\mathscr{I}$ the set of internal states and $\mathscr{E}$ the set of exiting states. The superscript d corresponds to the index Hierarchy (vertical location) in the tree, with $d \in \{1, ..., \mathscr{D}\}$ and subscript i is its horizontal location.

The tree structure is obtained by imposing an internal state at the root (level $d = 1$)[9]. This initial state then have several descendants which can be internal or emitting states. These sub-internal states can themselves have descendants which can be internal or emitting and so on. The transition from level d to level $d+1$ is provided by the probabilities of vertical transitions through an internal state. The return from level $d+1$ to level d is done with the exiting states and corresponds to a probability of exiting. Internal and emitting states of the same sub-HMM of level d communicate with a transition matrix as in a classical HMM model. An internal state leading to another internal/emitting states of a lowest level is called *parent* state. A parent state leads to *child* states. Finally, three probabilities govern the dynamic of the hidden structure:

- $\mathscr{A}_k^d = (a_k^d(i,j))$: is the matrix of horizontal transition of state i to state j where i and j are two states of the same sub-HMM of level d, i.e. $a_k^d(i,j) = \mathbb{P}[q_{t+1}^d = j | q_t^d = i]$, $q_{t+1}^d, q_t^d \in \{\mathscr{S}, \mathscr{I}\}$.
- $e_i^d = \mathbb{P}[q_{t+1}^d | q_t^{d+1}]$, with $q_t^{d+1}, q_{t+1}^d \in \{\mathscr{S}, \mathscr{I}\}$: vertical probability to leads from a child at level $d+1$ state to parent state at level d.
- $\pi_i^d = \mathbb{P}[q_{t+1}^{d+1} | q_t^d]$ with $q_t^{d+1}, q_{t+1}^d \in \{\mathscr{S}, \mathscr{I}\}$: vertical probability to lead from a parent state to its child state.

To ensure the transition matrix to be stochastic[10], it is necessary to impose that, for each sub-model depending of parent k :

$$\sum_{j\in ch(k)} a_k^d(i,j) + e_i^d = 1 \text{ and } \sum_{i\in ch(k)} \pi_i^d = 1 \tag{3.1}$$

[9] We will see later that this condition is not always necessary. The HRSDC model has not an internal state at the summit.

[10] $\sum_j a_{ij} = 1$ et $0 \leq a_{ij} \leq 1$.

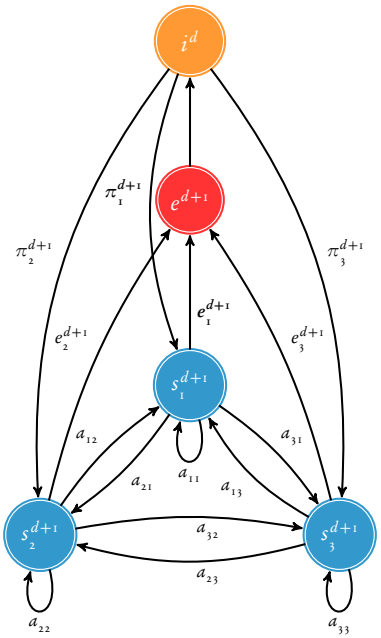

Figure 4: Basic structure of a HHMM with three emitting states.

where $i, j \in ch(k)$ are two states with parents k. Figure 4 is an illustration of the basic structure of a HHMM and figure 13 is a complete example (see appendix A). As we can seen, the number of parameters of the model depends on the structure of the tree. Xie (2005) uses a useful notation to represent parameters space of a HHMM. Let $\underbrace{Q}$ the size of the space state of each sub-model, Xie assume that global configuration of a HHMM starting from the root up to the d^{th} level can be written as:

$$q^{(d)} = \overline{(q^1 q^2 ... q^d)} = \sum_{i=1}^{d} q^i \underbrace{Q}^{d-i}$$

Thus, assuming that there is only one state at the root, all the parameters comprising a HHMM matches to:

$$\theta = (\cup_{d=2}^{D} \cup_{i=0}^{\underbrace{Q}^{d-1}-1} \{\mathcal{A}_k^d, \pi_i^d, e_i^d\}) \cup (\cup_{i=0}^{\underbrace{Q}^{D-1}} \{\phi_i\})$$

where ϕ_i corresponds to the parameters of a parametric probability distribution.

3.1.2 The Hierarchical RSDC model

The objective of the model we are building is to vary the correlations between two extreme major regimes, while allowing the existence of secondaries regimes. Like Silvennoinen and Teräsvirta (2007), the correlation process is bounded by four states of constant correlations over time. The structure highlights two primaries regimes, depending on abstract states i_1^1 and i_2^2. Each of these abstract states is connected with emitting states. Thus, the regime corresponding to i_1^1 is determinated by the emitting states s_1^2 and s_2^2; that of i_2^1 by s_3^2 et s_4^2. Figure 5 shows the hierarchical hidden structure of the Hierarchical RSDC model (HRSDC).

The hierarchical structure allows states to increase the granularity of the regimes. It establishes different types of regimes, which in our case are primaries and secondaries. The primaries regimes correspond to the regimes obtained with a classical Markov Switching model. To a higher level of granularity, these primaries regimes are built with sub-regimes, known as secondaries regimes. The structure allows secondaries's to capture nuances of dynamics that are thinner than the primaries's. The idea of granularity is illustrated on the figure 6. Our model has two levels and four secondaries regimes, corresponding to two sub-HMM models.

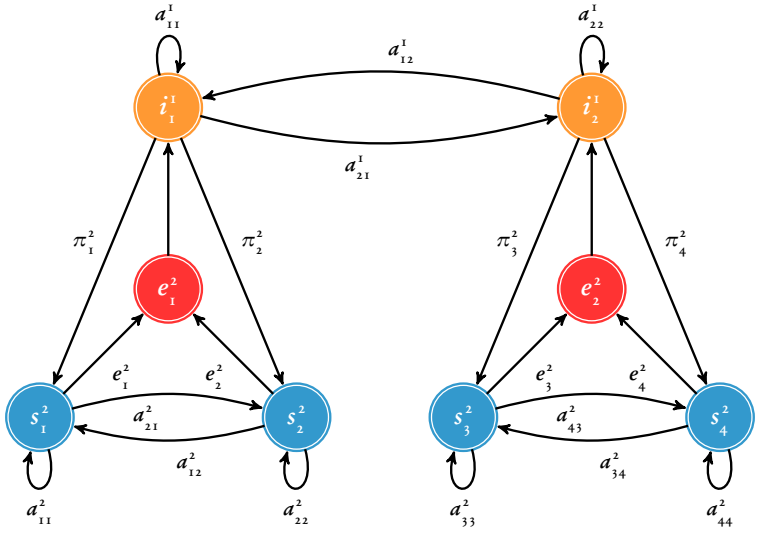

Figure 5: Hierarchical Hidden structure of the HRSDC.

The pair of emitting states defined by (s_1^2, s_2^2) forms a Markov-Switching model and the same is true for (s_3^2, s_4^2). The link between these sub-models is provided by the abstract states i_1^1 and i_2^1. The model is then building on two sub-models with two emitting states each, which transition matrix are respectively[11] :

$$A_1^2 = \begin{bmatrix} a_{11}^2 & a_{12}^2 \\ a_{21}^2 & a_{22}^2 \end{bmatrix} \text{ and } A_2^2 = \begin{bmatrix} a_{33}^2 & a_{34}^2 \\ a_{43}^2 & a_{44}^2 \end{bmatrix}$$

and verified constraints:

$$\begin{cases} a_{11}^2 + a_{21}^2 + e_1^2 = 1 \\ a_{12}^2 + a_{22}^2 + e_2^2 = 1 \end{cases} \text{ and } \begin{cases} a_{33}^2 + a_{43}^2 + e_3^2 = 1 \\ a_{34}^2 + a_{44}^2 + e_4^2 = 1 \end{cases}$$

where e_i^2, $i = 1, ..., 4$ is the probability of exiting from a state of level two and go to a parent state at level one. The two sub-HMM communicate via exiting states through abstract states i_1^1 and i_2^1. The dynamic in the transition from one to another of these abstract states is defined by the transition matrix:

$$A^1 = \begin{bmatrix} a_{11}^1 & a_{21}^1 \\ a_{12}^1 & a_{22}^1 \end{bmatrix}$$

which verifies:

$$a_{11}^1 + a_{12}^1 = 1 \text{ and } a_{21}^1 + a_{22}^1 = 1$$

Parameters π_i^2, $i = 1, ..., 4$ represents the probability to move from a parent state of first level to one of its children at the second level. These probabilities must verify:

$$\pi_1^2 + \pi_2^2 = 1 \text{ and } \pi_3^2 + \pi_4^2 = 1$$

The specification for the four correlation matrix constants in time is that outlined by Pelletier (see equations 2.29 and 2.31)[12]. In fact, the only difference with the RSDC is the hierarchical hidden structure which introduce us to see the RSDC as a special case of the HRSDC with only one level. As in the RSDC model, the specification defined by equation 2.29 can be estimate by EM algorithm whereas formulation defined by equation 2.31 allows to use iterative methods like Gradient.

[11] Recall that the probability that the state q which was in i at time $t-1$ to be in j at time t is written $\mathbb{P}[q_t = j | q_{t-1} = i] = p_{ji}$.

[12] The four correlation matrix then represent $4K(K+1)/2$ parameters.

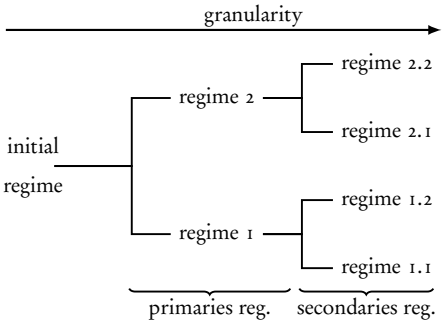

Figure 6: The increase of the granularity in the HRSDC model.

3.2 Estimation

Estimation of the HRSDC model can be done using the multi-step estimation of Engle and Sheppard (2001) and Engle (2002). This computationally attractive method splits up the log-likelihood $\ell(\varphi,\theta)$ as the sum of two parts : the volatility component $\ell_v(\varphi)$ and the correlations term $\ell_c(\varphi,\theta)$, where φ is the parameters space of the K univariate GARCH model and θ corresponding of the parameters of the correlations. More formally, this is written as:

$$\ell(\varphi,\theta)=\ell_v(\varphi)+\ell_c(\varphi,\theta)$$

with :

- $\ell_v(\varphi|\mathbf{r}_t)=-\frac{1}{2}\sum_{t=1}^{T}(K\log(2\pi)+2\log(|D_t|)+\mathbf{r}_t' D_t^{-2}\mathbf{r}_t)$
- $\ell_c(\theta|\mathbf{r}_t,\varphi)=-\frac{1}{2}\sum_{t=1}^{T}(\log(|R_t|)+\epsilon_t' R_t^{-1}\epsilon_t)$

Estimation of the volatility part ℓ_v is done by maximizing the sum of the individual GARCH likelihoods. Estimation of the correlation part is trickier because of the abstract states. Various solutions have been proposed.

In their seminal article, Fine *et al.* (1998) used a *generalized Baum-Welch* algorithm based on an modified version of the *Inside-Outside* algorithm. This method requires to calculate all sequences that could be generated by each sub-model of each level of the hierarchy. With a computational complexity of $\mathcal{O}(NT^3)$, where N represents the number of hidden states and T the number of observations, this approach is suitable for small sequences but remains impractical for long sequences[13]. An alternative approach, first developed in Paskin and Murphy (2001) -and taking up again but with some modifications by Bui *et al.* (2004)- provides an estimate by $\mathcal{O}(T)$. The idea is to consider the HHMM as a special case of Dynamic Bayesian Network (DBN). Note that a mechanic of a DBN is to repeat a chain of Bayesian network for each observation. The chains of the network are connected in time by a causal relationship which permit to track the evolution of the process. After the HHMM converted into a DBN, the estimation uses an classical Baum-Welch algorithm. However, if this is attractive from a computational point of view, it is still relatively complicated to implement. A third approach, suggested by Wierstra (2005), is to circumvent the problem posed by vertical transitions, which implies the existence of several paths to go from one state to another, by transforming the HHMM into a equivalent HMM. This *flattening* method, has the advantage of allowing to use the usual estimation methods of HMM with a complexity of $\mathcal{O}(T)$. Wierstra's tip is to redistribute vertical and horizontal probabilities such that no longer exists self-referential probabilities for internal states. This transformation allows to have a flat version of the HHMM in which there is only one path from a state to another. Xie (2005) succeeds to estimate the HHMM with HMM's standard tools while respecting the vertical dynamic of the model. In contrast to previous approaches, Xie's is builded on a particular expression of the transition matrix. Instead of considering the whole dynamic of the model, the transition matrix is broken down into several sub-transition matrix for each hierarchical level. Each level of the tree is then linked with a transition matrix. This rewriting of the transition matrix allows to estimate the correlations with a version of the Baum-Welch algorithm very similar to that used in the context of standard HMM. The complexity of Xie's approach is of order $\mathcal{O}(DTN^{2N})$.

[13] The inability of the Fine *et al.*'s algorithm to deal with long sequence is not presented as a problem by the authors they apply their model for unsupervised handwriting recognition. In such applications, the size of the sequences are about fifty characters.

To estimate our HRSDC model, we follows a modified Xie's approach. The main advantage of its method is the speed of implementation. Xie builds the transition matrix by successive layers in order to have an expression of the likelihood of complete data that can be factorizable. In this framework, vertical transitions, which allows to activate the child of a parent state are given by:

$$\pi_q = \prod_{d=1}^{D} \pi^d_{q^d},\ q = 0,...,\underbrace{Q^D - 1} \tag{3.2}$$

Transition matrix is rewritten by layers, each layer represents a level of the hierarchy of the tree:

$$\tilde{a}^d(q',q) = \prod_{i=d}^{D} e^i_{q'^i} \pi^i_{q^i} \cdot a(q'^d, q^d) \tag{3.3}$$

This probabilities are called *hypertransition probabilities*. Aggregation of these probabilities leads to the *hypertransition matrix* of our model:

$$\tilde{A}(q',q) = \sum_{i=1}^{D} \tilde{a}^i(q',q) \tag{3.4}$$

Finally, the hypertransition matrix has size $(S \times S)$ where $S = card(\mathscr{S})$. This formulation is attractive in the sense that it leads to work with a transition matrix similar to the classical HMM or Markov-Switching model. The hypertransition matrix of our HRSDC model is then written as:

$$\tilde{A} = \begin{bmatrix} a^2_{11} + e^2_1 \pi^2_1 a^1_{11} & a^2_{21} + e^2_2 \pi^2_1 a^1_{11} & e^2_3 \pi^2_1 a^1_{12} & e^2_4 \pi^2_1 a^1_{12} \\ a^2_{21} + e^2_1 \pi^2_2 a^1_{11} & a^2_{22} + e^2_2 \pi^2_2 a^1_{11} & e^2_3 \pi^2_2 a^1_{12} & e^2_4 \pi^2_2 a^1_{12} \\ e^2_1 \pi^2_3 a^1_{21} & e^2_2 \pi^2_3 a^1_{21} & a^2_{33} + e^2_3 \pi^2_3 a^1_{22} & a^2_{43} + e^2_4 \pi^2_3 a^1_{22} \\ e^2_1 \pi^2_4 a^1_{21} & e^2_2 \pi^2_4 a^1_{21} & a^2_{34} + e^2_3 \pi^2_4 a^1_{22} & a^2_{44} + e^2_4 \pi^2_4 a^1_{22} \end{bmatrix} \tag{3.5}$$

The hidden hierarchical structure requires twenty parameters. However, in practice, only sixteen parameters will be to estimate due to constraints of stochastic matrix. In the next sub-section, we propose two ways to estimate the correlations. The first is based on the EM algorithm. The second is done with the Hamilton's filter.

3.2.1 Estimation by EM algorithm

To run EM algorithm[14], we need to write the quantity $\mathscr{Q}(\theta|\theta_k)$. With our hypertransition matrix, it is simply written as:

$$\begin{aligned} \mathscr{Q}(\theta|\theta_p) &= \mathbb{E}_{\theta_p}[\log f(\epsilon_{1:T}, s_{1:T}; \theta)|\epsilon_{1:T}] \\ &= \sum_{t=1}^{T}\sum_{i=1}^{S} \log(f(\epsilon_t; \phi_i))\mathbb{P}_{\theta_p}[s_t = i|\epsilon_{1:T}] \\ &\quad + \sum_{t=1}^{T-1}\sum_{d=1}^{D}\sum_{i=1}^{S}\sum_{j=1}^{S} \log(\tilde{a}^d_{ij})\mathbb{P}_{\theta_p}[s_{t+1} = j, s_t = i|\epsilon_{1:T}] \\ &\quad + \sum_{i=1}^{S} \log(\nu_i)\mathbb{P}_{\theta_p}[s_1 = i|\epsilon_{1:T}] \end{aligned} \tag{3.6}$$

To simplify the later formula 3.6, we define:

$$\begin{aligned} \gamma^p_t(q) &\overset{def}{=} \mathbb{P}_{\theta_p}[q_t = q|\epsilon_{1:T}] \\ \xi^p_t(q',q,d) &\overset{def}{=} \mathbb{P}_{\theta_p}[q_t = q', q_{t+1} = q, e_t = d|\epsilon_{1:T}] \end{aligned}$$

Reestimation formulas are obtained by maximizing the expected value of the complete-data log-likelihood $\mathscr{Q}(\theta|\theta_p)$. In order to speed up this maximization, we can rewriting $\gamma^p_t(q)$ and $\xi^p_t(q',q,d)$ with the auxiliary variables defined by the Forward-Backward algorithm.

[14] See appendix B for a very brief review.

Forward-Backward algorithm The Forward-Backward algorithm is a useful method to reduce the calculus complexity[15]. It is to define two auxiliary quantities from which will be expressed reestimation formulas of the M step.

- Forward algorithm: the auxiliary quantity Forward, written $\alpha_t(q)$, is defined as the probability that the sequence $(\epsilon_1,...,\epsilon_t)$ to be in state q in the model θ :

$$\alpha_t(q) \overset{def}{=} \mathbb{P}_\theta[\epsilon_{1:t}, q_t = q]$$

Let $f^p(\epsilon_t;\phi_q)$ the value of the likelihood at time t and iteration p. To simplify the notations, we will omit the superscript p and use the short-hand notation of $f_q(y_t)$. The two steps calculus for Forward quantity is as follow:

$$\begin{array}{l} ① \text{ Initialization: } \alpha_1(q) = \pi_q f_q(\epsilon_1) \\ ② \text{ Iteration: } \alpha_{t+1}(q) = f_q(\epsilon_{t+1}) \sum_{q'} \sum_d \alpha_t(q_t) \tilde{a}^d(q',q) \end{array}$$

for $t = 1,...,T-1$ and $q = 0,...,\underbrace{Q^D}-1$.

- Backward Algorithm: the Backward variable, written $\beta_t(q)$, is the inverse of the Forward. It is defined as the probability to generate the sequence $(\epsilon_{t+1},...,\epsilon_T)$ by leaving from state q in the model θ:

$$\beta_t(q) \overset{def}{=} \mathbb{P}_\theta[\epsilon_{t+1:T} | q_t = q]$$

We have then two steps to compute the Backward variable:

$$\begin{array}{l} ① \text{ Initialization: } \beta_T(q) = 1 \\ ② \text{ Iteration: } \beta_t(q) = \sum_{q_t} \sum_d \beta_{t+1}(q') f_{q'}(\epsilon_{t+1}) \tilde{a}^d(q,q') \end{array}$$

for $t = 1,...,T-1$ and $q = 0,...,\underbrace{Q^D}-1$.

Auxiliary variables $\alpha_t(q)$ and $\beta_t(q)$ allows to rewriting the probabilities $\gamma_t^p(q)$ and $\xi_t^p(q',q,d)$ as follow[16]:

$$\begin{array}{l} \gamma_t^p(q) = \alpha_t(k)\beta_t(k) \\ \xi_t^p(q',q,d) = \alpha_t(q')\tilde{a}^d(q',q) f_{q'}(\epsilon_{t+1})\beta_{t+1}(q) \end{array}$$

As these two variables are probabilities[17], they shall normalize as:

$$\begin{array}{l} \bullet \ \sum_k \gamma_t^p(q) = 1 \\ \bullet \ \sum_k \sum_{k'} \sum_d \xi_t^p(q',q,d) = 1 \\ \bullet \ \gamma_t^p(k') = \sum_k \sum_d \xi_t^p(k',k,d) \end{array}$$

Then, we can easily rewriting the reestimation formulas with $\gamma_t^p(q)$ and $\xi_t^p(q',q,d)$.

Reestimation formulas Reestimation formulas can be obtained by maximizing $\mathcal{Q}(\theta|\theta_p)$ subject to the stochastic constraints on the parameters (see sub-section 3.1.2). Then, we have only to differentiate with respect to the constraints and adding the appropriate Lagrange factors. We take the Xie's notations to indexing the states. The writing $q = \overline{(rir')}$ and $q' = \overline{(rir'')}$ are two states configuration which are identical up to the d^{th} level and have the same $(d-1)^{th}$ parent r such that $r = q^{1:d-1} = q'^{1:d-1}$. Moreover, $r' = q_t^{d+1:D}$ and $r'' = q_{t+1}^{d+1:D}$ are two state configurations of a level below d. For each level

[15] it lets you go of a complexity from $\mathcal{O}(e^T)$ to $\mathcal{O}(T)$ (see Cappé (2004) chap. 3.2 and also Baum *et al.* 1966 p.170).

[16] For a detailed calculation, see Cappé (2004, chap. 3).

[17] In practice, the calculation of auxiliary variables Forward and Backward revealed problems of *underflow*. It is therefore advisable to apply a scaling factor in order to facilitate the computation (see appendix C).

$d \in \{1,...,\mathscr{D}\}$, the reestimation variables at iteration $p+1$ can be displayed as:

- $\hat{\pi}_r^d(j) = \dfrac{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_i \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)}{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_j\sum_i \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)}$

- $\hat{e}_r^d(i) = \dfrac{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_{q'}\sum_{d'\leq d} \xi_t^p(\overline{(rir')},q',d')}{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\gamma_t^p((rir'))}$

- $\hat{a}_r^d(i,j) = \dfrac{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''} \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)}{\sum_{t=1}^{T-1}\sum_{r'}\sum_{r''}\sum_j \xi_t^p(\overline{(rir')},\overline{(rjr'')},d)} \times (1-\hat{e}_r^d(i))$

- $\hat{R}_q = \dfrac{\sum_{t=1}^{T}\epsilon_t \cdot \gamma_t^p(q)\cdot \epsilon_t'}{\sum_{t=1}^{T}\gamma_t^p(q)}$

We simply use as stopping rule the conventional difference $||\hat{\theta}^{p+1} - \hat{\theta}^p|| \leq 10^{-6}$.

3.2.2 Estimation Gradient methods

Using Gradient methods to estimate a HMM is made possible with the so-called Hamilton's filter (see Hamilton (1989), (1990) and (1994, chap. 22)). This iterative filter allows to make inference on the state of the unobserved Markov chain. With Hamilton's notations, let $\hat{\xi}_{t|t}$ a vector of size $(N \times 1)$ which elements $\xi_{jt} = \Pr[s_t = j|\mathscr{F}_{t-1},\theta]$, $j = 1,...,N$ are the conditional probabilities to be in regime j given the information set $\mathscr{F}_{t-1}$ at time $t-1$ and $\tilde{A}$ our hypertransition matrix of size $(N \times N)$ (see equation 3.4). Then, the vector η_t of size $(N \times 1)$, which elements are the densities under the N regimes, i.e. $f(\epsilon_t|s_t = j,\mathscr{F}_{t-1};\theta)$, $j = 1,...,N$. Then, Hamilton shows that the observed density ϵ_t given the information set is written as:

$$f(\epsilon_t|\mathscr{F}_{t-1};\theta) = 1'(\hat{\xi}_{t|t-1} \odot \eta_t) \tag{3.7}$$

where $1'$ is a $(N \times 1)$ vector with all elements equal to 1. The conditional distribution of s_t can be deduced :

$$\begin{aligned} \frac{f(\epsilon_t, s_t = j|\mathscr{F}_{t-1};\theta)}{f(\epsilon_t|\mathscr{F}_{t-1};\theta)} &= \mathbb{P}_\theta[s_t = j|\epsilon_t,\mathscr{F}_{t-1}] \\ &= \mathbb{P}_\theta[s_t = j|\mathscr{F}_t] \end{aligned} \tag{3.8}$$

Combining with equation 3.7, relation 3.8 has the following expression :

$$\mathbb{P}_\theta[s_t = j|\mathscr{F}_t] = \frac{f(\epsilon_t, s_t = j|\mathscr{F}_{t-1};\theta)}{1'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{3.9}$$

As the joint density $f(\epsilon_t, s_t = j|\mathscr{F}_{t-1};\theta)$ is written as:

$$f(\epsilon_t, s_t = j|\mathscr{F}_{t-1};\theta) = \mathbb{P}_\theta[s_t = j|\mathscr{F}_{t-1}] \times f(\epsilon_t|s_t = j,\mathscr{F}_{t-1};\theta) \tag{3.10}$$

which is the density of the j^{th} regime. By collecting for the N regimes, the filtered probabilities become:

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{1'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{3.11}$$

with forecasts calculated by:

$$\hat{\xi}_{t|t+1} = \tilde{A} \times \hat{\xi}_{t|t} \tag{3.12}$$

Inference for each date t is then found by iterating equations 3.11 and 3.12. For choosing starting value, we use one of the options proposed by Hamilton (1994, p.693). He suggests to take the vector of unconditional probabilities π, which is the solution of the following system:

$$\begin{cases} \tilde{A}\pi = \pi \\ \mathbf{1}'\pi = 1 \end{cases} \tag{3.13}$$

The starting value are then given by:

$$\pi = (M'M)^{-1}Me_{N+1} \text{with } M = \begin{bmatrix} \mathbf{I}_N - \tilde{A} \\ \mathbf{1}' \end{bmatrix}$$

where e_{N+1} is the $(N+1)^{th}$ column of $\mathbf{I}_{N+1}$.

3.2.3 Smoothed probabilities

Hamilton's filter allows to make inference about the state of the Markov chain at time t conditional on the information set up to time t. Kim (1994) has developed an filter in order to make inference with the whole information set. Instead of computing $\xi_{t|t}$, it permit to compute $\xi_{t|T}$ with $t < T$. The Kim's filter is computed with our hypertransition as in the classical case:

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \{\tilde{A}'[\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}]\} \tag{3.14}$$

where $\oslash$ denotes element-by-element division.

4 Applications

This section contains Monte-Carlo experiments and two applications based on real dada. The first real database is that used by Colacito and Engle (2006) and contains daily data from S&P500 and 10-year bond futures (see appendix E). The second real database application is done with that of Pelletier (2006). It contains four week-days close exchange rates against US dollar which are Pound, Deutschmark, Yen and Swiss-Franc (see appendix D).

4.1 Simulated data

In this sub-section, we compare the correlations estimates of our HRSDC model and the DCC_{ES} of Engle and Sheppard (2001) in a setting where the true correlation structure is known. For simplicity, it is done in a bivariate framework. We simulate six DGP. The first one has the following variance equations:

$$h_{1,t} = 0.01 + 0.05r_{1,t-1} + 0.85h_{1,t-1}$$
$$h_{1,t} = 0.12 + 0.1r_{1,t-1} + 0.85h_{1,t-1}$$

and the others (DGP2:5):

$$h_{1,t} = 0.01 + 0.04r_{1,t-1} + 0.95h_{1,t-1}$$
$$h_{1,t} = 0.01 + 0.2r_{1,t-1} + 0.5h_{1,t-1}$$

For the DGP1, the correlations process follow a TAR-CCC model with constants correlations. The DGP2:5 are builded with dynamic processes. The different DGP are labeled as:

- DGP1: $R_{12,t} = \begin{cases} 0.99 \text{ if } t \in [1;250] \\ -0.99 \text{ if } t \in [251;500] \\ 0.4 \text{ if } t \in [501;750] \\ -0.4 \text{ if } t \in [751;1000] \end{cases}$
- DGP2: $R_{12,t} = 0.4 + 0.34\cos(t/250)$
- DGP3: $R_{12,t} = 0.5 + \cos(d/s) - (1/3)\cos(3d) + (1/7)\cos(5d)$, $d = (t-50)/145$, $s = 35$

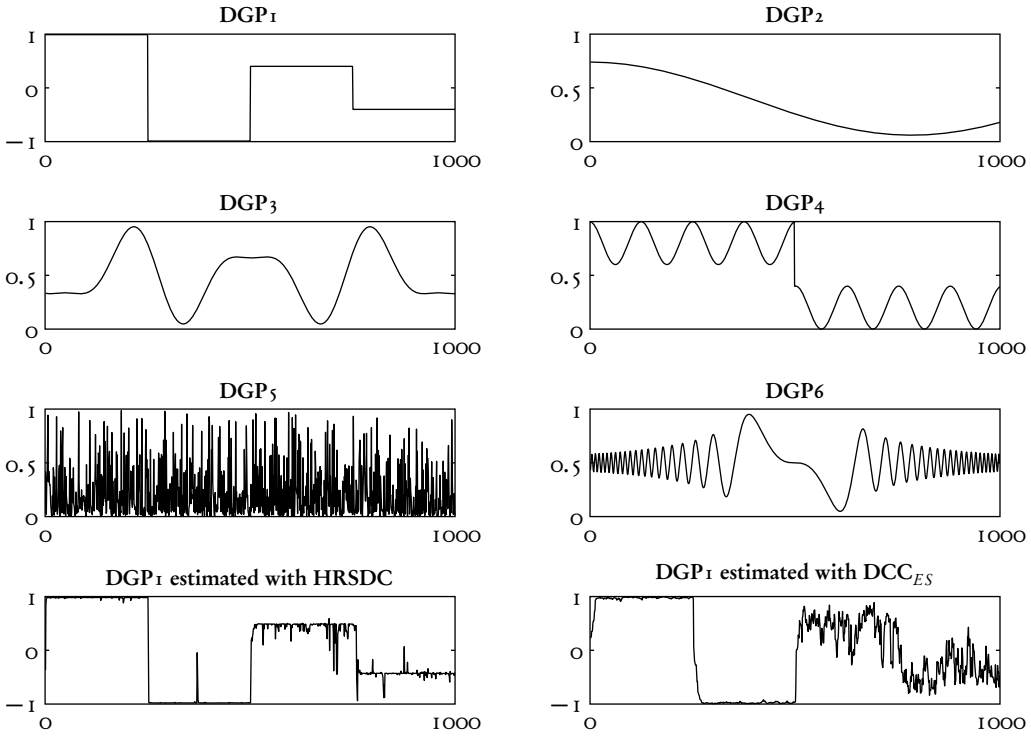

Figure 7: Simulated Correlation Processes. The last two graphs are an example of the estimated correlations of the two models for the DGP1.

- DGP4: $R_{12,t} = \begin{cases} .8 + .2\cos(t/20) \text{ if } t \in [1;500] \\ 0.2 + .2\cos(t/20) \text{ if } t \in [501;1000] \end{cases}$
- DGP5: $R_{12,t} = 0.99 - \frac{1.98}{1+\exp(0.5\max(\epsilon^2_{1,t-1},\epsilon^2_{2,t-1}))}$
- DGP6: $R_{12,t} = 0.5 + \sin(s^3)/(1+\sqrt{|s^3|})$, $s = 5 - t/100$

The simulated correlations are voluntarily pathological in order to test the accuracy of the correlations estimates corresponding of a very volatile/distress periods of financial markets. The DGP2:4;6 are builded with sinusoidal functions to create correlations with different regimes or sub-regimes. The DGP5 is used by Long and Ullah (2005) and corresponds of the stylized fact pointed out by Longin and Solnik (2001) that correlations among assets tend to increase during volatile periods.

The performance measures we use are very similar of theses used by Engle (2002). We first calculate two versions of a very classical loss function, which are computed as follow:

$$MAE_1 = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{R}_t - R_t\right|$$
$$MAE_2 = \frac{1}{T}\sum_{t=1}^{T}(\hat{R}_t - R_t)^2$$

For the second type of measure, we follow the methodology of Engle (2002) by considering the loss function of the Value-at-Risk (VaR). Recall that for a portfolio with a share w invested in the in the first asset and $(1-w)$ in the second, the VaR assuming normality can be computed as:

$$VaR_t^{\alpha} = \Phi_t^{-1}(\alpha)\sqrt{(w^2\hat{H}_{11,t} + (1-w)^2\hat{H}_{22,t} + 2w(1-w)\hat{R}_{12,t}\sqrt{\hat{H}_{11,t}\hat{H}_{22,t}})}$$

The loss function of the VaR is then defined by:

$$hit_t = \mathbb{1}_{\{wr_{1,t}+(1-w)r_{2,t}<-VaR_t\}} - \alpha$$

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
MAE$_1$						
DCC	.1331	.5125	.3774	.4237	.5298	.2764
HRSDC	.0562	.5521	.3836	.4642	.5386	.2831
MAE$_2$						
DCC	.0553	.2901	.1700	.2361	.2824	.0778
HRSDC	.0235	.3408	.1660	.2659	.3008	.0835
DQ à 5% *(strategy EW)*						
DCC	.7594	.5884	.6499	.9299	.5784	.9586
HRSDC	.5966	.6985	.7814	.9699	.4263	.9800
DQ à 1% *(strategy EW)*						
DCC	.3359	.9944	.9970	$1e^{-4}$	.0217	.9925
HRSDC	.4382	.9733	.9995	$2e^{-4}$	.0215	.9929
DQ à 5% *(strategy LS)*						
DCC	.7969	.1616	.5136	.9532	.7284	.7068
HRSDC	.8731	.0912	.3509	.9216	.6574	.7321
DQ à 1% *(strategy LS)*						
DCC	.4294	$4.3e^{-5}$	.9986	.9996	.01	.0523
HRSDC	.3896	.2685	.9959	.9994	.01	.0136

Table 1: Performance measures results.

We then use the *in-sample Dynamic Quantile* (DQ) test introduced by Engle and Manganelli (2001). This consists to test if all the coefficients of the regression of the violation process hit_t with its lagged values and others exogenous variables are equals to zero. To compute this F test, we use as explanatory variables five lags hit_t and the current value of the VaR. We test two numbers of rejections (1% and 5%) and two portfolios: an equal-weighted ($w = 0.5$, strategy EW) and a long-short (1 and -1, strategy LS). Table 1 presents the results of the performance measures.

The results show that the HRSDC performs better than the DCC when correlations are piecewise constants. This result is consistent with what could be expected since the HRSDC varied correlations between several constants correlations matrix over time. Once the simulated correlations are no longer piecewise constants, the DCC model has better MAE, which, once again, seems normal. However, differences in the MAE between the DCC and the HRSDC are rather weak. The dynamic process resulting from a combination of regimes of the HRSDC seems to be not too bad face to the autoregressive dynamic of the DCC.

The results of the DQ test are more mixed and appear highly dependents on the strategy chosen. The DCC is the best for EW strategy, while HRSDC dominates on the LS strategy. A simple addition of the best values shows that the HRSDC seems preferable to DCC. Nevertheless, it would be unfair to conclude to the dominance of HRSDC under the DCC. It is just better to point out that the HRSDC remains credible face to the DCC.

Finally, the HRSDC remains a formidable efficiency when correlations are piecewise constants (DGP1), but becomes less effective when it is no longer the case (DGP2:6). It seems that the autoregressive dynamic of the DCC stay very efficient when correlations oscillate around a trend value. However, the interest of the HRSDC is in its explanatory power. Beyond a simple measure of correlations, its structure allows to bring out the components of the overall dynamic by increasing the granularity of regimes.

4.2 Correlations between S&P500 index futures and 10-year bond futures

The first application with a real database is based on the bivariate sample of Colacito and Engle (2006). It contains daily returns of S&P500 index futures and 10-year bond futures from January 1990 to August 2003 (see appendix E). This data are also used by Silvennoinen et Teräsvirta (2007) to apply their DSTCC model. The individuals volatilities are obtained by running a GARCH(1,1) model. We later obtain the correlations by running three different models in order to make comparisons. These models are HRSDC, DCC_{ES} and DSTCC. For this sample, the correlations of the DCC_{ES} and DSTCC are estimated using gradient methods while the correlations of the HRSDC are computed with EM algorithm. The figure 8

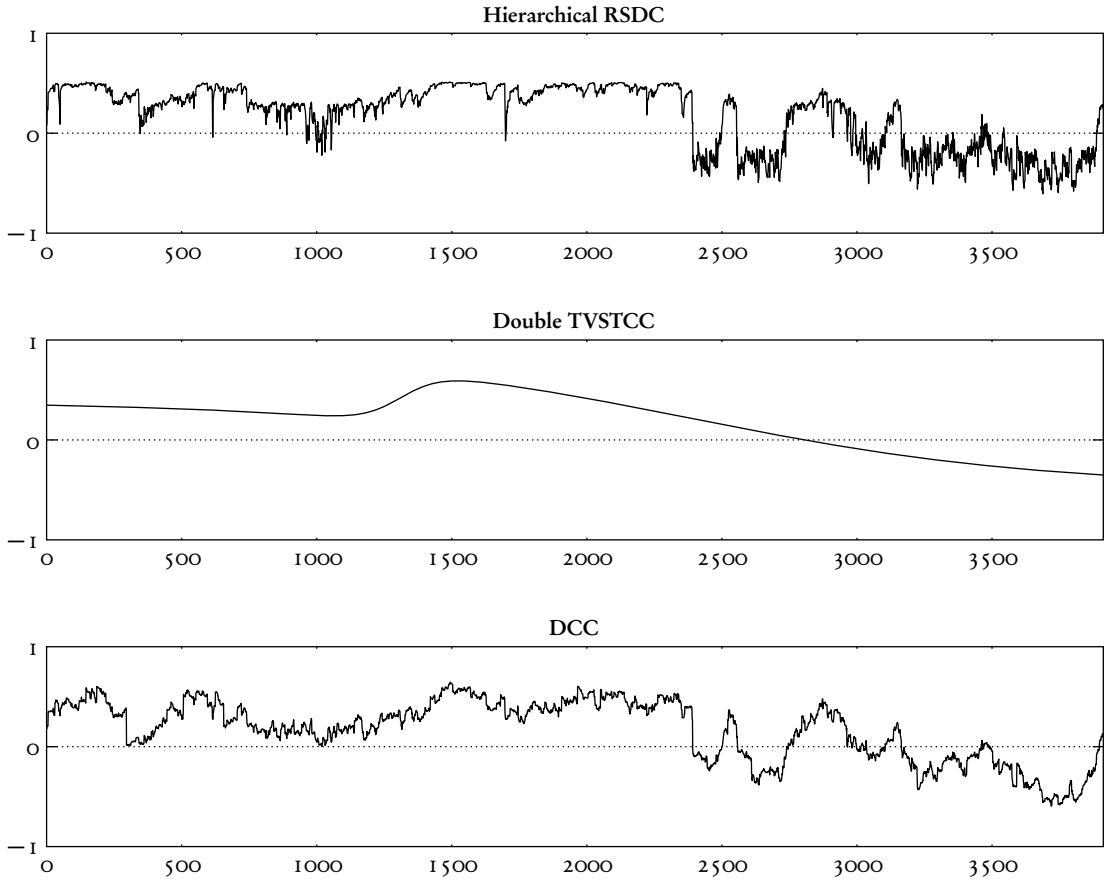

Figure 8: Estimated correlations of the HRSDC, DCC_{ES} and DSTCC with the sample of Colacito and Engle (2006).

shows the estimated correlations for the three models.

In our application, transition variables for the DSTCC are defined as $s_{1t} = s_{2t} = t/T$ (calendar time). In their study, Silvennoinen et Teräsvirta (2007) used another specification: one of the transition variable is calendar time while the other is VIX index[18]. This particular approach is motivated by investigate the fact that correlation among S&P500 and bond futures increase with distress of the market. As our model can not incorporate exogenous variable, calendar time allows to the DSTCC to be similar to the HRSDC. They are together conditional on time.

Results of the estimated parameters of the HRSDC are presented in appendix F and the smoothed probabilities corresponding to the secondary regimes are in the figure 9. The model has correctly identified two sub-regimes with positive correlations and two sub-regimes with negative correlations. Adding the smoothed probabilities of each pairs of sub-regimes give the smoothed probabilities that we would have had with Markov-Switching model with two regimes. This is shown on the Figure 10. The results for the DSTCC (with the calendar time specification) show that it does not accurately captures the *punctual* changes in regime. Correlations estimate have the form of a curve. Rather than a local measure, this model allows for a trend. For example, in the DCC and HRSDC models, correlations go to a negative correlations regime around the 2400^{th} observation whereas they are still positives in the DSTCC. These is due to the fact that the DSTCC has only two transition functions. A solution to this problem could be to introduce more transition functions as in Amado and Teräsvirta (2008). If the DCC and HRSDC correlations estimate seem similar, the advantage of HRSDC in the case study is to explain nuances in the dynamics through the decomposition in sub-regimes. The increasing of the granularity ends up to a

[18]The Vix index (for Chicago Board Options Exchange volatility index) is build with of selected basket of the implied volatility of S&P500 index options. It gives a measure of expectation of the market's volatility over the next 30 days. The VIX index can be freely downloaded at: `http://www.cboe.com/micro/vix/introduction.aspx`

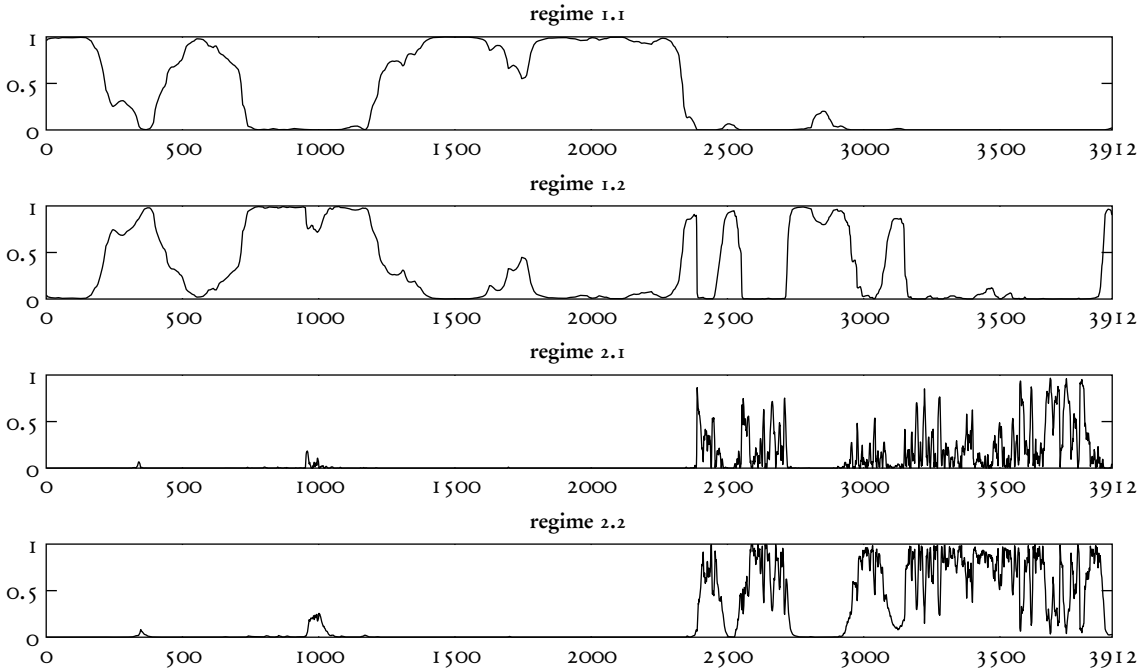

Figure 9: Smoothed probabilities of the HRSDC. Regimes 1.1 and 1.2 correspond to the emitting states $\{s_1^2, s_2^2\}$ and regimes 2.1 and 2.2 to $\{s_3^2, s_4^2\}$.

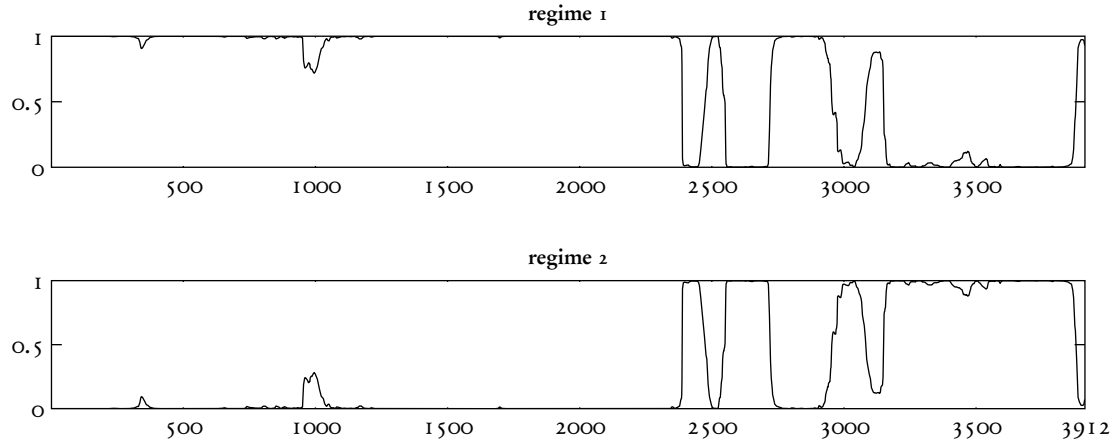

Figure 10: Smooth probabilities of the regimes defined by the internal states i_1^1 and i_2^1 (first application).

finer definition of the dynamic.

4.3 Correlations between exchange rate data

In this second application, we apply our HRSDC model to the sample used by Pelletier (2006). This series are exchange rate data and are plotted on the figure 14 (see also appendix D). As before, the dynamic of the standard deviations is obtained with a GARCH(1,1) model. For this case study, only DSTCC and HRSDC models are considered. We use the two-step maximum likelihood estimation using gradient methods for both models. Figure 11 shows plots of the estimated correlations for the two models. Parameters estimated for the HRSDC are in the table 3 (see appendix F). Smooth probabilities for the HRSDC can be seen in the figure 12. As we can seen on figure 11, as in the previous application, the DSTCC (with its calendar time specification, i.e. $s_{1t} = s_{2t} = t/T$) allows mainly to establish a trend on

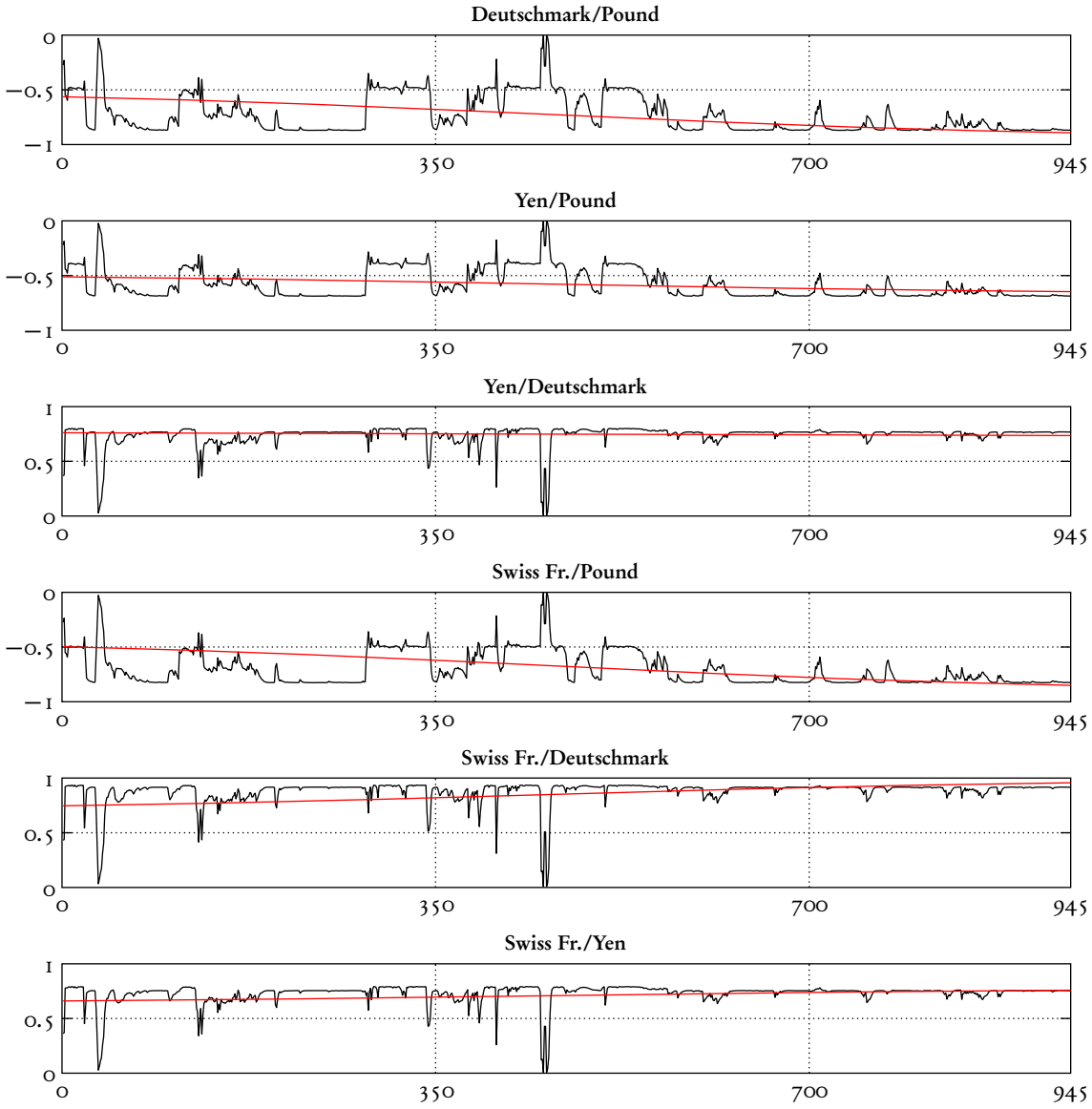

Figure 11: Estimated correlations for the HRSDC (in black) and DSTCC (in red).

the evolution of correlations in the sample. The reason is often that, with only two transition functions, the DSTCC can not capture more than three regimes. The direct consequence of this formulation is that it fails to capture the punctual extreme regimes, as it shows by example on the Swiss Fr/Deutschmark correlations. These are indeed marked by two peaks of correlations close to zero as they oscillate the rest of the time around a value about 0.8.

On this sample, the HRSDC clearly identifies four sub-regimes. Recall that the first primary regime associated with the internal state i_1^1 is a combination of the secondaries regimes depending of the two emitting states s_1^2 and s_2^2. Increasing the granularity provided by the hidden hierarchical structure allows to highlight the existence of a sub-regime linked to s_2^2. This sub-regime occurs only very rarely along the sample, around the 450^{th} observation. But it permit to capture an extreme and punctual behavior of the correlations process. This element of the global dynamic of correlations would be go unnoticed with a classical Markov-Switching model (because too short and not significant enough in relation to the size of the sample).

This application with a sample of four series brings to light a problem in the correlations specifi-

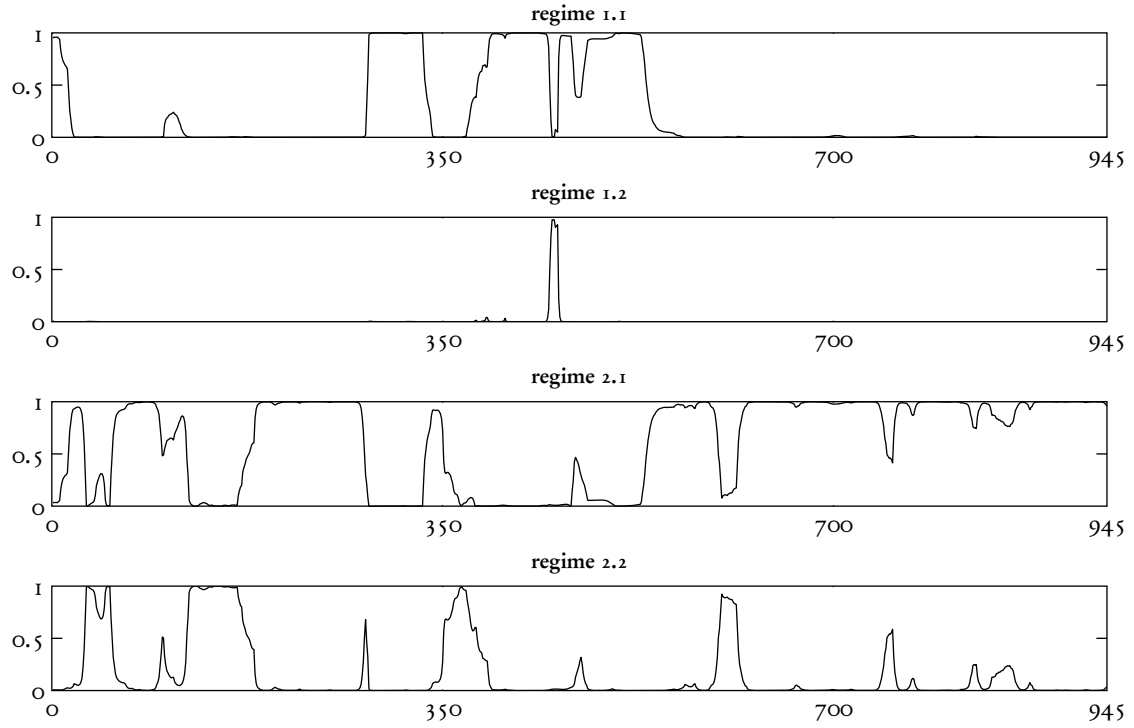

Figure 12: Smoothed probabilities for the HRSDC (exchange rate data).

cation. As we have said before, estimated parameters have been obtained using Gradient methods to estimate the HRSDC. This choice is not fortuitous. In fact, our experiments shows that the specification defined by equation 2.29 has some difficulties to return the maximum of the objective function by EM algorithm. The reason comes from the constraint Choleski representation used for the correlation matrix. Without this transformation, the derivatives of the objective function gets a formulation of the correlation matrix with no guarantees to have all one in the diagonal and PSD. But, the substitute of the Choleski representation for the diagonal elements:

$$c_{j,j} = \sqrt{r_{i,i} - \sum_{i=1}^{j-1} c_{j,i}^2} \tag{4.1}$$

by its constraint version:

$$c_{j,j} = \sqrt{1 - \sum_{i=1}^{j-1} c_{j,i}^2} \tag{4.2}$$

strongly disrupts the convergence of the EM algorithm. That's why it is preferable to use the specification defined by equation 2.31 which can be estimated by gradient methods using Hamilton's filter. In this example, despite of a huge number of iterations, the EM algorithm has not converge whereas Gradient methods rapidly reach the optimum of the objective function. Nevertheless, even the second specification (equation 2.31) need numerous parameters and the estimation of large correlation matrix could turn out to be cumbersome.

5 Conclusion

In this paper, we have presented a new multivariate GARCH with dynamic correlations. This further generalization, called Hierarchical RSDC (HRSDC), extends the RSDC model of Pelletier (2006). The HRSDC is a Markov-Switching class model with a correlation process bounded by four correlation

matrix constants over time. The innovation of our model is that it is builded on a hierarchical hidden structure introduced by Fine *et al.* (1998).

The main advantage of this hidden tree-like structure is to increase the granularity of the regime. It permits to define different types of regime; in our case, primary and secondary regime. Apply to correlations modeling, the HRSDC allows to capture thinner nuances than with the classical Markov-Switching approach. Monte carlo experiments and applications on real data show that this approach could improve understanding of the dynamic of the correlations. The application of the HRSDC to estimate the correlations between S&P500 index futures and 10-year bond futures or between exchange rate data has exhibited existence of sub-regime which can't appear with the another regime switching model.

While the results in this paper show that the HRSDC has a good explanatory power, several limitations have yet to be overcome. The first of them is about model selection. In this study, our model is builded from a symmetric hidden tree, with two primaries regimes which are formed with one sub-HMM each. This is a very simple structure. It is possible to build an asymmetric tree with much levels of depth. But finding the best hierarchy, i.e. split/merge/swap levels, is yet an open problem. Xie (2005) use the Reversible-Jump MCMC (RJMCMC) of Green (1997) for model selection. This approach seems interesting but it must be expanding. A second problem is about the specification of the correlation matrix. The specification of Pelletier doesn't suit to model large correlation matrix. As the hierarchical hidden structure model is a plug-in method, it could be interesting to find a specification for large correlation matrix. These tasks await future research.

References

[1] C. Amado, T. Teräsvirta. *Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure*. SSE/EFI Working Paper Series in Economics and Finance No. 691, 2008.

[2] F. Audrino, F. Trojani. *Estimating and Predicting Multivariate Volatility Thresholds in Global Stock Markets*. Journal of Applied Econometrics 21(3):345-369, 2006.

[3] L.E. Baum and T. Petrie. *Statistical Inference for Probabilistic Functions of Finite State Markov Chains*. The Annals of Mathematical Statistics, 37:1554-1563, 1966.

[4] L.E. Baum, T. Petrie, G. Soules, and N.Weiss. *A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains*. The Annals of Mathematical Statistics, 41:164-171, 1970.

[5] L. Bauwens, S. Laurent, J.V.K. Rombouts. *Multivariate GARCH Models : a Survey*. Journal of Applied Econometrics, 21(1)79-109, 2006.

[6] R.P. Berben, W.J. Jansen. *Comovement in International Equity Markets: a Sectoral View*. Journal of International Money and Finance, 24:832-857, 2005.

[7] P. Billingsley. *Probability and Measure*. 2^{nd} edition, Wiley Series in Probability and Mathematical Statistics, Wiley-Interscience, New-York, 1986.

[8] M. Billio, M. Caporin, M. Gobbo. *Flexible dynamic conditional correlation multivariate GARCH for asset allocation*. GRETA Working Paper, No. 04.03, 2004.

[9] D. Bohning, W. Seidel, M. Alfo, B. Garel, V. Patilea, G. Walther. *Advances in Mixture Models*. Computational Statistics & Data Analysis 51, 5205-5210, 2007.

[10] T. Bollerslev. *Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model*. Review of Economics and Statistics, 72:498-505, 1990.

[11] L. Cappiello, R.F. Engle, K. Sheppard. *Asymetric Dynamics in the Correlations of Global Equity and Bond Returns*. Journal of Financial Econometrics, 4(4):537-572, 2006.

[12] O. Cappe. *Ten years of HMMs*. Available from: `http://www.tsi.enst.fr/~cappe/docs/hmmbib.pdf`, 2001.

[13] O. Cappe, E. Moulines, T. Ryden. *Inference in Hidden Markov Models*. Springer-Verlag, Berlin, 2005.

[14] R. Colacito, R.F. Engle. *Testing and Valuing Dynamic Correlations for Asset Allocation*. Journal of Business and Economic Statistics, 24:238-253, 2006.

[15] L. Bauwens, A. Preminger, J. Rombouts. *Regime switching GARCH models*. CORE DP 2006/11.

[16] H.H. Bui, D.Q. Phung, S. Venkatesh. *Hierarchical Hidden Markov Models with General State Hierarchy*. In D. L. McGuinness and G. Ferguson, editors, Proceedings of the Nineteenth National Conference on Artificial Intelligence, 324-329, San Jose, California, USA, AAAI Press / The MIT Press, 2004.

[17] A.P. Dempster, N.M. Laird, D. B. Rubin. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society Series B, 39(1):1-38, 1977.

[18] R.F. Engle, K. Sheppard. *Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH*. Working Paper 2001-15, UCSD, 2001.

[19] R.F. Engle. *Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models*. Journal of Business and Economic Statistics, 20:339-350, 2002.

[20] Y. Feng. *A Local Dynamic Conditional Correlation Model*. MPRA Paper No. 1592, 2007.

[21] S. Fine, Y. Singer, N. Tishby. *The Hierarchical Hidden Markov Model : Analysis and Application*s. Machine Learning, 32(1):41-62, 1998.

[22] J.Mc. Geoffrey, K. Thriyambakam. *The EM Algorithm and Extensions*. Wiley Series in Probability and Mathematical Statistics, Wiley-Interscience, New-York, 1997.

[23] L. Glosten, R. Jagannathan, D. Runkle. *Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks*. Journal of Finance, 48:1779-1801, 1993.

[24] G. Gonzales-Rivera. *Smooth Transition GARCH Models*. Studies in Nonlinear Dynamics and Econometrics, 3:61-78, 1998.

[25] C. Franq, M. Roussignol, J.M. Zakoian. *Conditional Heteroskedasticity Driven by Hidden Markov Chains*. Journal of Time Series Analysis 22(2):197-220, 2001.

[26] P.H. Franses, D. van Dijk. *Non-linear Time Series Models in Empirical Finance*. Cambridge University Press, 2004.

[27] S. Gray. *Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process*. Journal of Financial Economics, 42:27-62, 1996.

[28] P. Green. *Reversible Jump MCMC Computation and Bayesian Model Determination*. Biometrika 82:711-732, 1995.

[29] C. Hafner, D. van Dijk, P. H. Franses. Semi-Parametric Modelling of Correlation Dynamics. In Advances in Econometrics, ed. by T. Fomby, C. Hill, and D. Terrell, vol. 20/A, pp. 59–103,. Amsterdam: Elsevier Sciences, 2005.

[30] G.E. Hagerud. *A Smooth Transition ARCH Model for Asset Returns*. Working Paper Series in Economics and Finance No. 162, Stockholm School of Economics, 1996.

[31] J.D. Hamilton. *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*. Econometrica, 57(2):357-384, 1989.

[32] J.D. Hamilton. *Analysis of Time Series Subject to Changes in Regime*. Journal of Econometrics, 45:39-70, 1990.

[33] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

[34] J.D. Hamilton, R. Sumsel. *Autoregressive Conditional Heteroskedasticity and Change in Regime*. Journal of Econometrics, 64, 307-333, 1994.

[35] M. Haas, S. Mittnik, M.S. Paollela. *A New Approach to Markov-Switching GARCH Models*. Journal of Financial Econometrics, 2(4):493-530, 2006.

[36] M. Haas, S. Mittnik, M.S. Paollela. *Mixed Normal Conditional Heteroskedasticity*. Journal of Financial Econometrics 2(2):211-250, 2004.

[37] C. Harvey, E. Ruiz, N. Shephard. *Multivariate Stochastic Variance Models*. Review of Economic Studies, 61:247-264, 1994.

[38] M. Lanne, P. Saikkonen. *Non-Linear GARCH Models for Highly Persistent Volatility*. The Econometrics Journal, 8(2):251-276, 2005.

[39] F. Longin, B. Solnik. *Extreme Correlation of International Equity Markets*. Journal of Finance , 56:651-678, 2001.

[40] F. Longin, B. Solnik. *The Asymptotic Distribution of Extreme Stock Market Returns*. Journal of Business, 69(3), 1996.

[41] F. Longin, B. Solnik. *Is the Correlation in International Equity Returns Constant : 1960-1990 ?*. Journal of International Money and Finance, 14(1), 1995.

[42] C. Li, W. Li. *On a Double-Threshold Autoregressive Heteroscedastic Time Series Model*. Journal of Applied Econometrics 11(3):253-274, 1996.

[43] X. Long, X, A. Ullah. *Nonparametric and Semiparametric Multivariate GARCH Model*. Unpublished manuscript, 2005.

[44] H.-M. Krolzig. *Markov-switching vector autoregressions, in: Modelling, Statistical Inference and Application to Business Cycle Analysis*. Lecture Notes in Economics and Mathematical Systems, vol. 454, Springer, Berlin, 1997.

[45] H.-M. Krolzig. *Constructing Turning Point Chronologies with Markov-Switching Vector Autoregressive Models: the Euro-Zone Business Cycle*. In: Eurostat (ed.), Proceedings on Modern Tools for Business Cycle Analysis, Monography in Official Statistics, 2003.

[46] C.J. Kim. *Dynamic Linear Models with Markov-Swithcing*. Journal of Econometrics, 60:1-22, 1994.

[47] G. Koutmos. *Asymmetries in the Conditional Mean and the Conditional Variance: Evidence from Nine Stock Markets*. Journal of Economics and Business 50(3):277-290, 1998.

[48] G.J. McLachlan, D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Mathematical Statistics, Wiley-Interscience, NewYork, 2000.

[49] H. Lee, J. Yoder. *A Bivariate Markov Regime Switching GARCH Approach to Estimate Time Varying Minimum Variance Hedge Ratios*. Applied Economics, forthcoming, 2006.

[50] B.G. Leroux. *Maximum-Likelihood Estimation for Hidden Markov Models*. Stochastic processes and their applications, 40 :127-143, 1992.

[51] G. Lindgren. *Markov Regime Models for Mixed Distributions and Switching Regressions*. Scandinavian Journal of Statistics, 5:81-91, 1978.

[52] S. Lundbergh, T. Teräsvirta. *Modelling Economic Highfrequency Time Series with STAR-STGARCH Models*. Working Paper, Stockholm School of Economics, 1999.

[53] J.C. Liu. *Stationarity of a Markov-Switching GARCH Model*. Journal of Financial Econometrics, 4(4):573-593, 2006.

[54] F. Klaassen. *Improving GARCH Volatility Forecasts with Regime-Switching GARCH*. Empirical Economics, 27, 363-394, 2002.

[55] K. Murphy, M. Paskin. *Linear Time Inference in Hierarchical HMMs*. Proceedings of Neural Information Processing Systems, 2001.

[56] J. Marcucci. *Forecasting Stock Market Volatility with Regime-Switching GARCH Models*. Studies in Nonlinear Dynamics and Econometrics, 9(4), article 6, 2005.

[57] M.C. Medeiros, A. Veiga. *Modeling Multiple Regimes in Financial Volatility with a Flexible Coefficient GARCH Model*. Econometric Theory, to appear, 2008.

[58] X.L. Meng, D. van Dyk 1997. *The EM Algorithm - An Old Folk-Song Sung to a Fast New Tune.* Journal of the Royal Statistical Society Series B, 59(3):511-567, 1997.

[59] K. Nam, C. Pyun, A. Arize. *Asymmetric Mean-Reversion and Contrarian Profits: Anst-garch Approach*. Journal of Empirical Finance 9(5):563-588, 2002.

[60] A. Palandri. *Sequential Conditional Correlations : Inference and Evaluation.* Workong Paper, Duke University, 2005.

[61] D. Pelletier. *Regime Switching for Dynamic Correlations*. Journal of Econometrics, 131(1-2):445-473, 2006.

[62] L.R. Rabiner. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*. Proceedings of the IEEE, 1989.

[63] R. Rabemananjara, J. Zakoian. *Threshold ARCH Models and Asymmetries in Volatility*. Journal of Applied Econometrics, 8(1):31-49, 1993.

[64] E. Sentana. *Quadratic Arch Models*. Review of Economic Studies 62(4):639-661, 1995.

[65] A.N. Shiryaev. *Probability*. 2nd edition, Springer-Verlag, Berlin, 1996.

[66] A. Silvennoinen, T. Teräsvirta. *Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations*. SSE/EFI Working Paper Series in Economics and Finance No. 577, 2005.

[67] A. Silvennoinen, T. Teräsvirta. *Modelling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model*. SSE/EFI Working Paper Series in Economics and Finance No. 652, 2007.

[68] A. Silvennoinen, T. Teräsvirta. *Multivariate GARCH models*. SSE/EFI Working Paper Series in Economics and Finance No. 669, 2007.

[69] T. Teräsvirta. *Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models*. Journal of the American Statistical Association, 89:208-218, 1994.

[70] H. Tong and K.S. Lim. *Threshold Autoregressive, Limit Cycles and Data*. Journal of the Royal Statistical Society B, 42:245-292, 1980.

[71] D. Smith. *Markov-Switching and Stochastic Volatility Diffusion Models of Short-Term Interest Rates*. Journal of Business and Economic Statistics 20(2):183-197, 2002.

[72] D. van Dijk, T. Teräsvirta, P.H. Franses. *Smooth Transition Autoregressive Models - A Survey of Recent Developments*. Econometric Reviews, 20(1):1-47, 2000.

[73] Y.K. Tse, A.K.C. Tsui. *A Multivariate GARCH Model with Time-Varying Correlations.* Journal of Business and Economic Statistics, 20:351-362, 2002.

[74] D. Wierstra. *Statistical Models for Non-Markovian Control Tasks*. Master's Thesis, Institute of Information and Computing Sciences, Utrecht University, 2004.

[75] C.F.J. Wu. *On the Convergene Properties of the EM Algorithm*. Annals of the Statistics, (11):95-103, 1983.

[76] L. Xie. *Unsupervised Pattern Discovery for Multimedia Sequences*. PhD Thesis, Columbia University, 2005.

A An example of HHMM

Figure 13 is an example of HHMM with four levels and six emitting states.

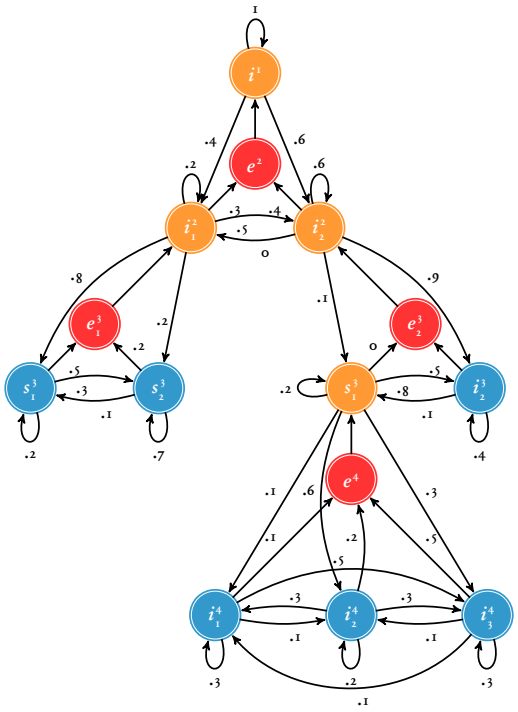

Figure 13: An example of Hierarchical Hidden structure.

B EM algorithm

The origin of the Expectation-Maximization (EM) algorithm comes from the difficulties to estimate the parameters of models based on unobserved latent variables. More specially, it is an alternate approach face to usual iterative method to reach the optimum of an objective function. In the case of HMM/Markov-Switching models, the difficulties comes generally from two levels: (i) the likelihood function has numerous optimums (ii) optimizing the likelihood function is analytically intractable.

The firsts major contributions to elaborate this technique are due to Baum *et al.* (1966 and 1970). They developed the two step methodology which consist to first compute the expectation of the likelihood function with the latent variables (E step) and then maximizing this expected likelihood (M step). From a theoretical point if view, the condition of convergence of the algorithm were proved by Dempster *et al.* (1977) and improved few years later by Wu (1983). The literature on the subject was later expanded considerably and many variants have emerged (see for example Meng and van Dyk (1997) and also Geoffrey and Thriyambakam (1997)).

More formally, the mechanic of the EM algorithm is to proceed by successive maximization based on observable data and information that we have every iterations on the latent variables. Instead of trying to directly calculate the conditional expectation $\mathbb{E}[\log f(\epsilon, s|\theta)|\epsilon, \theta]$, which is impossible because of s, the problem is being circumvented by focusing on the calculation of the quantity:

$$\mathcal{Q}(\theta|\theta_p) \overset{def}{=} \mathbb{E}_{\theta_p}[\log f(\epsilon_{1:T}, s_{1:T}; \theta')|\epsilon_{1:T}]$$

that it is calculable. Because the log-likelihood of the complete data is convex, it is possible to use Jensen inequality to deduce that, for all (θ', θ) : $\mathcal{Q}(\theta'|\theta) \geq \mathcal{Q}(\theta|\theta) \Longrightarrow \ell(\theta') \geq \ell(\theta)$. The two steps methodology for the p^{th} is then as follow:

$$\textcircled{1}\ \textit{E-step} : \mathcal{Q}(\theta|\theta_p) = \mathbb{E}_{\theta_p}[\log f(\epsilon, s; \theta')|\epsilon]$$
$$\textcircled{2}\ \textit{M-step} : \hat{\theta}_{p+1} = \arg\max_{\theta} \mathcal{Q}(\theta|\theta_p)$$

By iterating ① and ②, the sequence $(\theta_p)_{p\in\mathbb{N}}$ ensures the convergence of the sequence $(\ell(\theta_p))_{p\in\mathbb{N}}$ to the true vector of parameters.

C Scaling factor

Computing the quantities $\alpha_t(q)$ and $\beta_t(q)$ can lead to underflow problem. A easy way to circumvent this issue is to introduce a *scaling factor*, written as ς. In that case, initialization of Forward variable, written $\breve{\alpha}_t(q)$ is done by taking $\varsigma(1)=\sum_q \alpha_1(q)$:

$$\breve{\alpha}_1(q)=\frac{\alpha_1(q)}{\varsigma(1)}$$

We can then compute the *intermediary Forward variable*, written $\bar{\alpha}_{t+1}(q)$:

$$\bar{\alpha}_{t+1}(q)=f_q(\epsilon_{t+1})\sum_{q'}\sum_d \breve{\alpha}_t(q_t)\tilde{a}^d(q',q)$$

which lead to the scaling factor $\varsigma(t+1)$:

$$\varsigma(t+1)=\sum_q \bar{\alpha}_{t+1}(q)$$

The adjusted Forward variable can then be written as follow:

$$\breve{\alpha}_{t+1}(q)=\frac{\bar{\alpha}_{t+1}(q)}{\varsigma(t+1)}$$

Similar calculus can be apply to obtain the Backward variable. By initializate with $\breve{\beta}_T(q)=1$, the intermediary quantity is expressed as:

$$\bar{\beta}_t(q)=\sum_{q_t}\sum_d \breve{\beta}_{t+1}(q')f_{q'}(\mathbf{y}_{t+1})\tilde{a}^d(q,q')$$

and the adjusted Backward variable is obtained by computing for each $t=T-1,...,1$:

$$\breve{\beta}_t(q)=\frac{\bar{\beta}_t(q)}{\varsigma(t)}$$

The two quantities $\breve{\alpha}_t(q)$ and $\breve{\beta}_t(q)$ allow to rewriting the probabilities $\gamma_t^p(q)$ and $\xi_t^p(q',q,d)$ as:

$$\gamma_t^p(q)=\breve{\alpha}_t(k)\breve{\beta}_t(k)$$
$$\xi_t^p(q',q,d)=\breve{\alpha}_t(q')\tilde{a}^d(q',q)f_{q'}(\mathbf{y}_{t+1})\breve{\beta}_{t+1}(q)$$

D Exchange rate data

The exchange rate database is that of Harvey *et al.* (1994) and was also used by Kim *et al.* (1998) and Pelletier (2006). It is freely downloadable at the url :

`http://www.nuffield.ox.ac.uk/users/shephard/pub.aspx`

It contains week-days close exchange rates against US dollar for Pound, Deutschmark, Yen and Swiss-Franc over the period 1/10/81 to 28/6/85. This series are plotted on figure 14.

E S&P500 index futures and 10-year bond futures

This database is that of Colacito and Engle (2006) (plotted on figure 15) and available at the url:

`http://www.unc.edu/~colacitr/Research/Files/EC2006.zip`

It contains daily returns of S&P500 index futures and 10-year bond futures from January 1990 to August 2003 (3912 observations).

F Estimated parameters of the HRSDC for the real database applications

Table 3 shows the parameters estimations of the correlations process for the database of exchange rate data; table 2 for the database of Engle and Colacito.

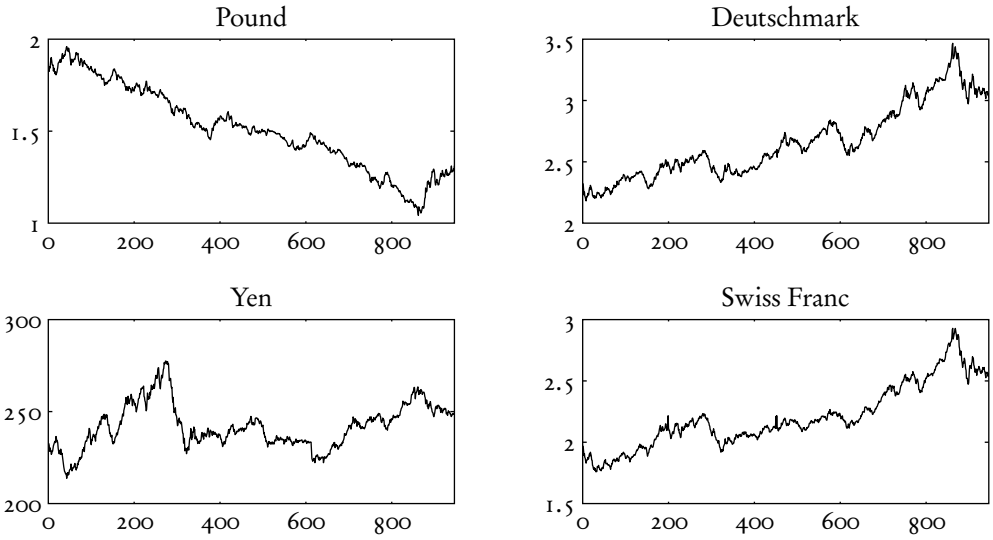

Figure 14: Exchange rate database.

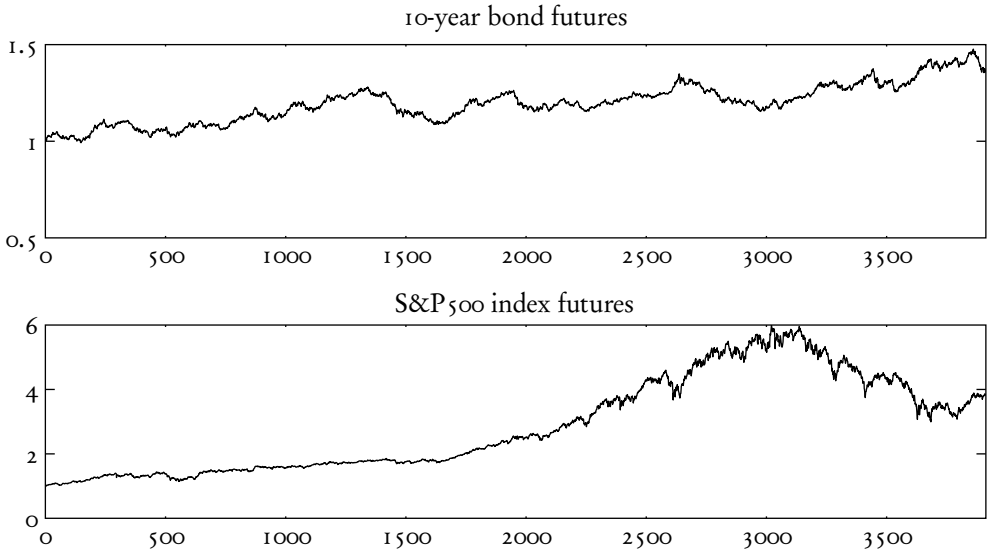

Figure 15: S&P500 index futures and 10-year bond futures.

correlation matrix:

$$R_1 = \begin{bmatrix} 1 & \\ 0,5189 & 1 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 1 & \\ 0.2924 & 1 \end{bmatrix}$$
$$R_3 = \begin{bmatrix} 1 & \\ -0.1853 & 1 \end{bmatrix} \qquad R_4 = \begin{bmatrix} 1 & \\ -0.1597 & 1 \end{bmatrix}$$

transition probabilities

$$A^1 = \begin{bmatrix} 0.9797 & 0.9365 \\ 0.0203 & 0.0635 \end{bmatrix}$$
$$A_1^2 = \begin{bmatrix} 0.9919 & 1.4e^{-4} \\ 0.0055 & 0.7876 \end{bmatrix} \; A_1^2 = \begin{bmatrix} 0.8139 & 0.0618 \\ 0.1853 & 0.9299 \end{bmatrix}$$
$$e_1^2 = 5.85e-4,\ e_2^2 = 0.2123,\ e_3^2 = 7.7e-4,\ e_4^2 = 0.0084$$
$$\pi_1^2 = 0.0067,\ \pi_2^2 = 0.9933,\ \pi_3^2 = 0.9933,\ \pi_4^2 = 0.0067$$

Table 2: Estimated parameters for the correlations of the second real data application.

correlation matrix:

$$R_1 = \begin{bmatrix} 1 & & & \\ -0,4860 & 1 & & \\ -0,3943 & 0,8125 & 1 & \\ -0,5004 & 0,9488 & 0,8032 & 1 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 1 & & & \\ -4e-5 & 1 & & \\ -3e-5 & 8e-05 & 1 & \\ -5e-5 & 9e-5 & 8e-5 & 1 \end{bmatrix}$$
$$R_3 = \begin{bmatrix} 1 & & & \\ -0,8746 & 1 & & \\ -0,6912 & 0,7720 & 1 & \\ -0,8299 & 0,9214 & 0,7587 & 1 \end{bmatrix} \qquad R_4 = \begin{bmatrix} 1 & & & \\ -0,7337 & 1 & & \\ -0,5798 & 0,6476 & 1 & \\ -0,6962 & 0,7730 & 0,6365 & 1 \end{bmatrix}$$

transition probabilities

$$A^1 = \begin{bmatrix} 0,9933 & 0,4167 \\ 0,0066 & 0,5832 \end{bmatrix}$$
$$A_1^2 = \begin{bmatrix} 0,6711 & 0,0084 \\ 0,0115 & 0,9915 \end{bmatrix} \; A_1^2 = \begin{bmatrix} 0,9835 & 0,0002 \\ 0,0157 & 0,9571 \end{bmatrix}$$
$$e_1^2 = 0.3172,\ e_2^2 = 8.21e-05,\ e_3^2 = 7.37e-04,\ e_4^2 = 0.0425$$
$$\pi_1^2 = 0.9933,\ \pi_2^2 = 0.0066,\ \pi_3^2 = 0.9933,\ \pi_4^2 = 0.0067$$

Table 3: Estimated parameters for the correlations of the second real data application.